Geographic Mobility of College Students and the Gender Gap in Academic Aspirations

Lídia Farré
University of Barcelona and IAE-CSIC

Francesc Ortega*
CUNY, Queens College

July 18, 2023

Abstract

We study the decision to pursue an advanced degree from an internationally renowned academic institution, which greatly facilitates access to top jobs. Relying on unique data on applications to a highly selective program that provides graduate fellowships to Spanish students, we show that women apply for the fellowships at lower rates than observationally equivalent male graduates in non-STEM fields, whereas the opposite is true in STEM fields. We also find that female students are relatively less interested in doctoral programs and less willing to study abroad than males in most fields of study. To shed light on the mechanisms, we surveyed current college students about their post-graduation plans. We confirmed that female college students are relatively less interested in doctoral studies and less geographically mobile than males, even after controlling for a wide range of observable characteristics. The lower geographic mobility of female students (in non-STEM fields) correlates with lower earnings expectations and greater involvement in romantic relationships.

JEL Classifications: J3, J7
Keywords: Gender; Post-graduate studies; Fellowships; Geographic Mobility

*We are grateful to the *La Caixa Foundation* and the participating universities for providing access to their data and for their overall support. We also thank Pol Campos-Mercadé, Rosa Ferrer, Paola Giuliano, Laura Hospido, Pedro Rey, Carlos Sanz and Judit Vall for helpful comments. We are particularly indebted to Amparo Asensi, Dolors Baena, David Casado, Domènec Espriu, Montserrat Masoliver, Jessica Pan, Pablo Pareja, Raul Ramos, Santi Roca and Josep Ros. We acknowledge the financial support of the Ministry of Science and Innovation (grant PID2019-110397RA-I00), the Government of Catalonia (grant SGR2017-644) and the *La Caixa Foundation*. E-mail: lidia.farre@ub.edu and francesc.ortega@qc.cuny.edu.

1 Introduction

In the last few decades gender gaps in the labor market have narrowed, largely thanks to the increase in women's educational attainment. Women account for over 60% of recent college graduates in many countries, but remain under-represented in top positions, both in academic settings and in the private sector (Blau and Kahn (2017)). A vast literature explores the many factors that contribute to this persistent gender gap.

Our paper focuses on a new angle: a college degree or even an advanced degree often do not guarantee access to top jobs. Both in academia and private sector, being hired for a high-paying position at a leading firm or public institution often requires a post-graduate degree from an internationally renowned university, particularly for entry-level positions. Despite accounting for the majority of college graduates, women remain in the minority in many prestigious graduate programs. Evidence from the field of Economics & Business illustrates this point. Top MBA programs remain majority male (Wallen et al. (2017)) and women account for only 32% of the entering cohorts in Economics Ph.Ds (Bayer and Rouse (2016), Boustan and Langan (2019), Beneito et al. (2021)).[1]

We investigate whether women make less (or more) ambitious post-graduate educational choices than their male counterparts, and whether the answer varies across fields of study. This could be a subtle, yet important, factor to help explain the absence of women in top positions in the labor market, impacting their chances to be hired for a top job or to remain in one when having children (as in Cortes et al. (2020)). More specifically, we examine the post-graduate plans and choices of college students through the lens of participation in a highly selective fellowship program. Gaining admission to internationally renowned graduate programs is difficult, because of the harsh competition. It is also expensive in terms of tuition and other expenses, and typically requires moving to a different city or country. For these reasons, and because of high social payoffs in terms of innovation and knowledge diffusion, governments and philanthropic institutions in many countries offer fellowships to academically excellent students interested in pursuing graduate studies at the world's leading universities.[2]

[1] Women are also under-represented in STEM fields that are typically associated with above-average labor market prospects, and are also less likely to pursue professional degrees and doctoral studies (Bertrand and Hallock (2001), Black et al. (2008), Hsieh et al. (2019)). Even within Economics, there are large gender disparities in the presence of women across subfields. Beneito et al. (2021) report substantial under-representation of women in Macroeconomics and Finance.

[2] One of the most famous graduate fellowship programs in the world is the *Fulbright U.S. Student Program*, established in 1946, and offering approximately 2,000 grants each year. Alumni of the program occupy leading positions across a wide range of professions.

We examine college graduates' decision to apply to the *La Caixa Foundation* (LCF) fellowship program, largely aimed at funding graduate studies abroad for Spanish citizens with excellent academic records. These data provide a unique window into the educational choices of high-achieving male and female college graduates. Specifically, we combine data on the whole set of applicants to the program over a number of years with administrative graduation records for four large universities that account for the majority of college students in the region of Catalonia. These data allow us to estimate the application rates of male and female college graduates by field of study. In addition, we were authorized to match individual records for the largest university in our data, which allows us to parse out the effects of academic ability and socioeconomic status, in addition to gender.

Our analysis of the administrative data clearly shows that GPA, socio-economic status and age at graduation are strong determinants of the decision to apply to the fellowship program with the expected signs. Namely, the probability of participating in the program increases rapidly with students' GPA and socio-economic status (parental education), and falls with age at graduation. In addition, female graduates in STEM fields (excluding Health & Life sciences) apply to the fellowship program at higher rates than men, despite being a minority in these fields. In contrast, female participation in non-STEM fields is substantially lower than men's, even after controlling for socio-economic status and age at graduation. Furthermore, we find that female students are relatively less interested in doctoral programs and less geographically mobile than males in all fields of study, with the exception of Health & Life sciences.

To gain a better understanding of post-graduation career plans and to shed light on the gender gaps in fellowship applications just described, we conducted a large survey among the (approximately 35,000) students currently enrolled in the four universities included in our study. Our analysis confirms that female college students are much less interested in doctoral studies, but moderately more interested in Master's degrees, than men even after conditioning on a rich set of individual characteristics. The survey analysis also indicates that female students (in non-STEM fields) tend to be less geographically mobile than similar males, and this is partly due to lower earnings expectations and greater involvement in romantic relationships. This finding relates to recent studies showing that commuting disproportionately penalizes women in the labor market (Le Barbanchon et al. (2019), Fluchtmann et al. (2020) and Petrongolo and Ronchi (2020)). Our findings imply that geographical distance is also a larger impediment for females in the context of their *educational* choices, introducing a *proximity-prestige* *tradeoff*.

Our results contribute to the literature on the absence of women in high-earnings, high-status positions, often referred to as the glass ceiling (Bertrand et al. (2019)). Several explanations have been proposed to account for the gender disparities at the top of the labor market. Early studies emphasized gender discrimination (Rouse and Goldin (2000)) and differences in skill levels (Goldin et al. (2006)). More recently, researchers have also documented gender differences in preferences for competition (Niederle and Vesterlund (2007), Buser et al. (2014), Hospido et al. (2019)) and in the balance between family and work (Bertrand (2013), Azmat and Ferrer (2017), Bursztyn et al. (2017), Keloharju et al. (2019) and Hospido et al. (2019)). In addition, several studies have also pointed out the role of reviewers in candidate selection processes, whose decisions may be affected by implicit bias, gender stereotypes or other factors (Bagues and Esteve-Volart (2010), Breda and Ly (2015), Hospido and Sanz (2019), Farré and Ortega (2021), Montalban and Sevilla (2020)). In particular, our findings suggest that explanations based on differences in human capital accumulation remain important: highly talented women in a variety of fields of study make less ambitious educational choices than their male counterparts. In line with previous studies, our results indicate that economic and family expectations are strongly correlated with future career plans (Azmat et al. (2020) and Wiswall and Zafar (2021)). However, we also show that *current* family ties, such as having a romantic partner, can also affect women's willingness to study abroad with likely negative effects on their careers.[3]

Last, our paper also contributes to the literature comparing the academic achievements of boys and girls. It is well established that, from an early age, girls "leave boys behind" in terms of educational attainment (Fortin et al. (2015)). However, the evidence is less clear in regards to the comparison between the most talented males and females. Our data contain individual records for over 160,000 college graduates and allow us to produce highly detailed comparisons of the GPA distributions by gender with high granularity. We find that while women typically have higher mean GPA than men, they tend to be under-represented in the top 5% of the grades distribution (for a given

[3]The issue of the geographical mobility of male and female workers has regained interest in the recent years. Several studies document that marriage, cohabitation and children reduce women's mobility to a larger extent than men's. For instance, Shauman and Xie (1996) find lower geographic mobility for female scientists. They argue that this is related to their higher likelihood of being in dual career marriages and that their mobility falls further, relative to their male partners, when they have children. Similarly, Jürges (2006) documents that marital status (and cohabitation) reduce the geographical mobility of women relative to men, confirming the earlier work of Bielby and Bielby (1992) and others.

major and university).[4] This difference could be relevant to explain the gender gaps in highly meritocratic contexts, such as seeking admission to graduate studies at a leading institution. Our estimates show that the probability to apply for a LCF fellowship is substantially larger for high-GPA students. However, the difference in the shares of men and women among high-grade earners are too small to explain away the observed gap in participation rates in the program.

The structure of the paper is as follows. Section 2 presents our data sources. Section 3 estimates aggregate application rates in the fellowship program using data for the four universities included in our study. Section 4 extends the analysis further by focusing on individually linked records (for a single university). Section 5 uses new survey data to examine the post-graduation career plans of college students and their geographic mobility. Section 6 concludes.

2 Data Sources

2.1 Applications to LCF Fellowships Program

The *La Caixa Foundation* (LCF) is a private financial institution in Spain that has been providing graduate fellowships since 1982. The LCF fellowship program is the largest program in Spain funding graduate studies abroad, currently awarding 120 fellowships per year (plus around 100 more for graduate studies in Spain).[5] The awards fund both Master's degrees and PhDs in all fields of study and the fellows typically gain admission to the most prestigious institutions worldwide.[6]

Our data contains applications for the period 2014-2018 to three separate LCF sub-programs: graduate studies in North America or Asia, in European countries (other than Spain) and doctoral studies in Spain.[7] The data contains complete information on

[4] The under-representation of women among extremely high-achieving students has also been documented for the United States among high school students in math (Ellison and Swanson (2010)). It is also well-known that males' aptitude test scores exhibit larger variance than females', and that males outnumber females among high-achievers along several (but not all) dimensions (Hedges and Nowell (1995)).

[5] Similar programs aimed at Spanish citizens are the Foundation Alonso Martin Escudero (60 fellowships), the Foundation Mutua Madrilena (40), Fulbright (25), the Ramon Areces Foundation (22), Rafael del Pino Fellowships (10) and the Barrie Foundation (10).

[6] To date, the LCF has funded close to 5,000 awards, with 70% funding studies abroad. The top destination countries are USA, Spain, UK, Germany and France. Similarly, the top (narrow) fields of study have been: Art & History (14%), Health and life sciences (13%), Engineering (13%) and Economics & Business (12%).

[7] The latter sub-program also requires geographic mobility. The host institution cannot be located

roughly 9,000 applications, 55% of which from female candidates. Besides gender, the data contains university of origin, graduation year, field of study, and the outcome of the selection process. The success rate (for complete applications) is around 9%.[8]

2.2 College graduates records

We obtained access to the individual (anonymized) graduation records of the 4 largest public universities in Catalonia: the University of Barcelona (UB), the Autonomous University of Barcelona (UAB), the Polytechnic University of Catalonia (UPC) and University Pompeu Fabra (UPF). These universities are located in the Barcelona metropolitan area and together account for 77% of the enrollment in public colleges in Catalonia.[9] Three of these universities offer a large number of majors across all fields of study, whereas UPC is almost completely specialized in engineering.

Our period of analysis ranges from academic year 2009-2010 to 2018-2019 and the data have wide coverage across all academic disciplines. For short, we refer to each graduation cohort on the basis of the Fall semester of the graduation year. Hence, following this convention, our data contains graduation cohorts for the period 2009-2018. Among the roughly 162,000 individual observations, about 43.3% of the graduates belong to Social Sciences, 31.2% to STEM disciplines (excluding Health and Life sciences), 13.7% to Health & Life sciences (including Biology, Biochemistry and Environmental sciences) and 11.8% to Arts & Humanities.[10]

The graduation records include student-level information on year of graduation, major, gender and GPA. The data show that 55.4% of all graduates are women. Across fields, we observe that they account for a large majority in all fields, except for STEM where women are only 28.8% of the graduates. The female share rises to 65.7% in Social Sciences, 67.5% in Arts & Humanities and 73.2% in Health & Life Sciences.

Graduation GPA is reported on a 0-10 scale (with a minimum of 5 required to pass

in the province where the candidate conducted his/her undergraduate studies.

[8]Women account for only 49% of the successful applicants, despite making up 55% of the applications. Farré and Ortega (2021) analyze the sources of this gender gap and show that it is due to differences in gender composition across fields of study combined with reviewers' preference for *gender balanced* outcomes *within* each field. Another study employing data from the *La Caixa Foundation* fellowship program is Garcia-Montalvo (2014), which showed that the labor market careers of award recipients experience a large and persistent boost (both in academia and private sector).

[9]Public colleges account for 85% of the overall (in-person) tertiary enrollment in Catalonia, which amounts to 173,485 students in academic year 2018-2019.

[10]Prior to academic year 2013-2014, the graduation records for the UAB are incomplete and do not include all their majors. We exclude the incomplete cohorts for this university from the analysis.

a class). Across all graduates, the mean GPA is 7.11. However, we observe differences by gender and also field of study. The average GPA for women is 7.23, about 4% higher than for men (6.96). By field of study, the highest mean GPAs are found in Health (7.45) and Arts & Humanities (7.42), followed by Social Sciences (7.16) and STEM (6.79).[11]

The University of Barcelona (UB) agreed to share with us information on students' age and family background (e.g. parental education and occupation), as well as to link their data with the LCF applications dataset at the individual level.[12] The UB is the largest university in our dataset, accounting for almost half of the graduates.[13] It is also fairly similar to the other universities in terms of the share of females (65% in academic year 2018-2019) and enrollment distribution across fields of study, with the exception of UPC that specializes in engineering and has a much smaller share of female students (about 30%). The LCF applications dataset contains 588 complete applications from UB graduates, corresponding to 506 unique individuals, over the period 2014-2018. The data show that 44 of these applicants were awarded a fellowship, that is, the success rate was 8.7%. This is almost exactly the overall success rate among all applicants to the LCF program (from all universities) in our data, which hovers around 9%.

2.3 Survey of post-graduation plans among college students

We conducted a survey of all students at our four participating universities that had completed over half of the 240 credit hours required for graduation (and had registered for at least one class in academic year 2019-2020). The survey was conducted online in January-February 2020 and the response rate was 14%, leading to 4,848 essentially completed questionnaires out of a target population of 34,559 students.[14]

Students received an e-mail invitation through their institutional account. Participation in the survey was incentivized with a chance to win one of three 100-euro vouchers

[11]Mean GPAs are very similar across all universities (ranging between 7.23 and 7.29) except for the engineering school where the mean value is 6.72.

[12]Obtaining permission to match administrative data across different sources has become much more difficult after the application of the General Data Protection Regulation (GDPR) in the European Union. This regulation was adopted in 2016 but implemented from May 25, 2018). To link the two datasets while preserving student anonymity, each party encrypted the students' National Identification Number using the same key. Then we simply merged the two datasets on the basis of the encrypted identifier.

[13]The overall number of UB graduates for academic years 2009-2010 through 2018-2019 is 75,596, or approximately 7,500 per year.

[14]This response rate is quite typical of online surveys conducted by these universities on their own student population. The response rate for our survey also compares favorably to Paredes et al. (2020). If we include questionaries that are only partially complete the response rate increases to 16%.

(to be spent at a popular bookstore) but was otherwise voluntary. The message clearly stated that the data collection was anonymous and that the information would be exclusively used for academic purposes. The invitation was signed by each university's vice-rector of research, the research team in this paper, and the LCF (which paid for the survey).

As we show in detail in Section 5, the survey respondents match fairly well the administrative records in terms of gender, field of study and GPA. In particular, 58% of the overall respondents are female, and the share of female students for each university in the survey data is very close to what we observe in the administrative records. In the survey, the share of female students ranges between 66% and 69% for the UB, UAB and UPF, falling to 30% for the engineering school (UPC). The corresponding values in the administrative records (graduating cohorts 2009-2018) are 60% to 64% and 27%, respectively.

Turning now to GPA, the mean value among all survey respondents is 7.2, only slightly above the average of 7.1 found in the administrative records. The average (self-reported) GPA in the survey ranges between 7.2 and 7.4 for the three universities with a broad range of majors but it is significantly lower at the engineering school (6.8). As we show later, these figures also match closely the corresponding numbers in the administrative records.[15]

3 Program participation rates

Let us now use our administrative data to compute the application rates to the LCF program, which we refer to as (program) participation rates, for each graduating cohort.

More specifically, we consider the set of all applicants to the LCF program (during calendar years 2014-2018) that graduated from the 4 universities participating in our study, and sort them by year of graduation. First, we compute the number of applicants in calendar year t from graduating cohort c ($Applicants_{c,t}$). We then construct the

[15] We also note that the students that complied and completed the survey tend to be positively self-selected in terms of grades: 32.6% and 18.8% have grades above the 75th and 90th percentiles, respectively (relative to the grade distribution based on university-field in the administrative records). The administrative records show that women, on average, obtain slightly higher grades than men in all universities (with female-male ratios ranging between 1.02 to 1.03) with the exception of the UPC, which effectively exhibits gender parity (0.996 FM ratio). The situation is similar in the survey data, although the gender gaps are narrower. For the UPC respondents, the female-male ratio is estimated to be 1.003, while the values for the other universities range between 1.001 and 1.014. The narrower GPA gender gaps are consistent with the higher response rates for women and high-GPA students.

program participation rate (PR) for cohort c in post-graduation year $t \geq c$ by dividing by the size of the cohort, that is,

$$PR_{c,t} = \frac{Applicants_{c,t}}{Graduates_c}, \tag{1}$$

where $Graduates_c$ is the number of students graduating in academic year c. Clearly, we can also compute participation rates at lower levels of aggregation, by restricting to applicants and graduates from a specific gender or field of study.[16]

It is worth noting that the likelihood that a given student participates in the LCF program upon graduation will be greatly influenced by a number of individual characteristics, such as GPA, age, or socio-economic status. In Section 4, we will estimate the roles of these characteristics for the graduates of one of the universities, which agreed to link the records of their graduates to the LCF applications, using an encryption algorithm to safeguard students' anonymity.[17]

3.1 Censoring

A natural yardstick to assess whether male and female students participate at similar rates in the LCF program is a cohort's *total* participation rate (TPR), which consists of the number of members of a cohort c that apply to the program in *any* year after graduation ($t \geq c$). Namely,

$$TPR_{c,t} = \sum_{t \geq c} PR_{c,t} = \frac{\sum_{t \geq c} Applicants_{c,t}}{Graduates_c}. \tag{2}$$

Obviously, an accurate estimation of the TPR requires data on applications to the LCF program over a long period of time. Unfortunately, we only have information on applications to the fellowship program over the period $t = 2014 - 2018$. This data limitation introduces a *censoring* problem in the estimation of the total participation rate and could also lead to biased estimates of the gender gap in TPR if the timing of participation decisions in the fellowship program varies by gender.[18]

[16] We can also compute participation rates by university. However, we were not authorized to report data for each university separately, except for the UB.

[17] The age and socio-economic status of applicants is not part of the LCF data, but university records include this information.

[18] To fix ideas, suppose 10 male and 10 female students graduate at $t = 0$ and suppose they can only apply to the LCF program one, two or three years after graduation ($t = 1, 2, 3$). Let us assume that females tend to apply early (i.e. shortly after graduation) whereas males delay their applications.

Our plan to mitigate potential censoring bias is to investigate the participation *age profile*, separately by gender, in order to identify which graduation cohorts suffer from the least degree of censoring.[19] Clearly, if participation in the fellowship program decreases in age (as we will show shortly), the cohort graduating in academic year 2013-2014 (referred to as $c = 2013$) will be the least affected by censoring bias. For this cohort, we are able to compute $PR(2013, t)$ for calendar years $t = 2014, ..., 2018$ and keep in mind that $PR(2013, 2013) = 0$ because the students in this cohort graduated in the second half of 2014 and did not qualify for the 2013 fellowship program.

Table 1 reports the participation rates for each cohort for each of the years for which we have LCF applications data (2014-2018), pooling both genders. As can be seen in column 3, for graduation cohort $c = 2013$, the total participation rate is $TPR = 1.87\%$. For this cohort, the highest (annual) participation rates were obtained in the year of graduation and one year later, 0.51% and 0.69%, respectively. Beyond that point in time, annual participation rates fall to much lower levels. The table also illustrates a substantial degree of variation in annual participation rates across cohorts. For instance, the highest annual PR is sometimes attained in the year of graduation (as was the case in cohorts 2014 and 2016) and sometimes one year later (as in cohorts 2013 and 2015). Nevertheless, in all cases the highest annual PRs are attained in the first two years and range between 0.43% and 0.69%. It is also important to recognize that a cohort's interest in graduate studies and, hence, the cohort's total participation rate in the LCF program will be affected by idiosyncratic factors, such as the state of the labor market in the years immediately after graduation. Furthermore, applying to the LCF program is a fairly rare event (with total participation rates below 2%). Thus, it would be risky to focus our analysis on a single cohort. Instead, we will consider additional cohorts in our analysis to mitigate variability associated with cohort shocks and increase overall sample size, while tolerating only small increases in censoring.

In order to decide which additional cohorts to consider, it is helpful to build the

Specifically, suppose that at $t = 1$, one woman applies but no men. At $t = 2$, 1 man and 1 woman apply. Last, at $t = 3$, only 1 man applies. Clearly, there is no gender gap in TPR since the value is 2/10 both for men and women. Suppose that the data for $t = 3$ is unavailable (censored). Based solely on periods $t = 1, 2$, the TPRs for females and males are 2/10 and 1/10, respectively. Thus, the female-male TPR ratio will be 2. Hence, if the age-participation profiles vary by gender, censoring will lead to a spurious gender gap. In this example, censoring leads to an overestimation of the total participation of females relative to males. It is worth noting that censoring problems of this nature are rather common in studies that require tracking cohorts over time.

[19] It would be more accurate to speak of *years-since-graduation* profiles because there is variation in the age within a graduation cohort. Unfortunately, we only know the age of the graduates for one of the universities in our data. We use the term *age profiles* because it is less cumbersome and more intuitive.

average *participation-age profile*. Namely, we define $PR(\tau)$ as the average $PR(c,t)$ across (c,t) pairs for which $\tau = t - (c+1) \geq 0$.[20] In particular, $PR(\tau = 0)$ averages $PR(2013, 2014)$, $PR(2014, 2015)$, $PR(2015, 2016)$, $PR(2016, 2017)$ and $PR(2017, 2018)$. Figure 1 plots the results for both genders pooled together. The resulting age profile starts off at a fairly high value (of 0.55%) in the year of graduation ($\tau = 0$) and one year later and falls almost monotonically thereafter, reaching a mere 0.02% eight years after graduation. Thus, missing data on applications shortly after graduation will severely underestimate the TPR. It is also worth pointing out that the precision of the average PR estimates will be the same for $0 \leq \tau \leq 4$ (as those estimates are all based on the average of 5 cohorts). For $\tau \geq 5$, the number of cohorts averaged falls gradually, until only a single cohort ($c = 2009$) is used in the estimation of the average PR for $\tau = 8$.

While estimation of the participation age-profiles is interesting in its own right, our main purpose is the estimation of gender *gaps* in participation. Clearly, if participation-age profiles vary by gender, censoring might bias our estimates of the gender gap in participation in the fellowship program. Thus, it is important to examine the participation age profiles of male and female students separately. As shown clearly in Figure 2, the profiles are very similar for male and female graduates. For both groups, PRs peak within one year of graduation and fall almost uniformly thereafter. The figure also suggests that the PRs of males are higher than females' in the year of graduation (and perhaps also 8 years after graduation).

Based on these results, we choose to pool 3 graduation cohorts: cohort $c = 2013$ and the adjacent cohorts $c = 2012, 2014$. Specifically, for the 2012 cohort, we observe LCF applications for 1-5 years after graduation (but lack the applications submitted on the year of graduation). For the 2013 cohort, we observe PRs for 0-4 years after graduation, and, for the 2014 cohort, we observe applications for 0-3 years since graduation. Thus, pooling these cohorts captures fairly well the data on applications for $\tau \leq 4$ years after graduation. As shown at the bottom of Table 1 (column 3), the TPR for this set of cohorts is only slightly lower than for the 2013 cohort (1.65% versus 1.87%) and much higher than if we had included all cohorts in our estimation (1.07%). Importantly, restricting the sample to the 2012-2014 cohorts is unlikely to introduce much bias in the estimation of the gender gap in TPR.[21]

[20] Recall that the cohort graduating in academic year 2013-2014 is denoted by $c = 2013$ and the earliest its members can apply to the LCF program is $t = 2014$.

[21] The resulting potential bias is likely to *overestimate* the female-male TPR ratio by a small amount. This lower-bound argument is based on the participation-age profiles estimated earlier. The aggregation of cohorts 2012, 2013 and 2014 entails three missing values: PR(2012,2013) (corresponding to

Next, we turn to the estimation of the gender gap in TPR by gender as well as by field of study.

3.2 Participation rates by gender and field of study

Having devised a plan to mitigate censoring bias, we can now turn to the estimation of the gender gap in total participation rates (TPR) by gender as well as by field of study. Table 2 collects our estimates. To set the stage, the top panel reports the TPR obtained when using all graduation cohorts ($c = 2009 - 2018$). The first column shows that the number of female graduates in our data is 23% larger than the number of male graduates. Not surprisingly, the number of female applicants to the LCF fellowship is also higher than the number of male applicants. However, the number of female applicants was only 15% higher, which implies a gender gap in total participation rates. Column 3 reports the TPR by gender over the period 2014-2018: for males, the value was 1.11%, whereas for women it was 1.04% (resulting in a female-male TPR ratio of 0.94).

As argued above, we do not observe applications in the years following graduation for our earlier cohorts, which severely bias the estimates for the TPRs, and could also affect the gender gap. To mitigate this problem, the second panel considers only graduation cohort $c = 2013$. As expected, TPRs increase substantially and the female-male TPR ratio also increases (to 1.09). However, the estimated TPRs by gender are based on fairly low numbers of applications (135 for male and 180 for females), which introduces high variability due to cohort-specific shocks when we focus on field-specific TPRs. The following two panels expand the sample by adding the adjacent cohorts. The third panel considers cohorts $c = 2013, 2014$ and the bottom panel includes $c = 2012$ as well. The same patterns emerge from both of these panels, which leads us to focus on the larger 3-cohort panel in order to improve the precision of our estimates. The data show that the overall TPRs were effectively equal for males and females in the bottom panel, at 1.65%. Hence, when we pool all fields of study, the estimates suggest that the participation rates

applications in the year of graduation for cohort 2012), PR(2013,2019) (corresponding to applications 5 years after graduation for cohort 2013), and PR(2014,2019) and PR(2014,2020) (corresponding to applications 4 and 5 years after graduation for cohort 2014). According to Figure 2, the participation rates of male and female students are practically equal 4 and 5 years after graduation. However, in the year of graduation, male students participate in the program at higher rates than female students. Thus, the male TPR estimated by pooling the 2012-2014 cohorts will probably be slightly downward biased. As a result, the female-male TPR ratio will be slightly biased upward. Furthermore, Figure 2 shows that the gender gap in PR in the year of graduation is roughly 0.10 percentage points. Because we pool 3 cohorts, the resulting bias in the estimated gender gap in TPR is expected to be a mere $0.10/3 = 0.03$ percentage points.

of male and female students are at parity.

However, the overall estimates mask important heterogeneity across fields of study, as shown in columns 4-7 in Table 2. To get a sense of the role of field of study, we grouped all majors into 4 broad areas: STEM, Health & Life sciences, Arts & Humanities and Social sciences. We shall focus on our preferred set of cohorts (2012-2014), but the pattern that emerges is similar across all panels in the table. Before examining gender gaps in participation, it is worth noting that TPRs vary substantially across fields (pooling both genders). As shown in columns 4-7 (bottom panel): the highest values are found in Health & Life sciences (3.45%) and in Arts & Humanities (2.65%) and the lowest in STEM (1.67%) and Social sciences (0.86%).[22] To the extent that female students specialize in different fields than men, we will observe differences in the overall TPR resulting from pooling all fields together. The bottom of the table reports the shares of students across fields, separately by gender. In our data, 50.3% of male students major in STEM fields, compared to only 16.6% among females. In addition, females are over-represented in high-TPR fields (Health & Life sciences and Arts & Humanities), but this is largely offset by their over-representation in the field of study with the lowest TPR (Social sciences), as evidenced by the parity in the female-male TPR ratio when pooling all fields.

Let us now examine the *within-field* gender gaps in TPR. Let us consider first the TPR by gender in STEM fields (column 4). According to our data, the TPRs for men and women are STEM is 1.59% and 1.87%, respectively. Thus, the TPR for female students is estimated to be 18% higher than for male students. While this estimate strongly suggest that females in STEM apply to the fellowship program at higher rates than men, it is worth noting that only 16.6% of all female graduates (in cohorts 2012-2014) majored in STEM fields, which makes the estimated TPR for women somewhat imprecise.[23] In stark contrast, columns 5-7 show that in non-STEM fields female TPRs are lower than the corresponding values for males. The estimated female-male TPR ratios are 0.88 in Health & Life sciences and in Arts & Humanities, and 0.80 in Social

[22]Many factors may explain why interest in graduate studies abroad differs across fields. An obvious one is that Social Sciences includes majors in Law and Social work that involve a great deal of country-specific knowledge and are less conducive to studying abroad.

[23]Only 87 females graduated in STEM in cohorts 2012-2014 and applied for the LCF fellowship (between 2014 and 2018). Our data also show that, for all four universities in our study, the STEM female-male TPR ratios are above one (ranging between 1.06 and 2.55), as shown in Table A.11. It is also worth noting that the gender-specific TPRs obtained when pooling all universities (or all fields) are weighted sums of the corresponding values for each of the universities (or fields). However, this is not the case for the female-male ratios.

sciences. Thus, females in non-STEM fields appear to participate in the LCF fellowship program at lower rates than men.

In sum, our analysis of the administrative data suggests that the relative participation of female students in the LCF program varies by field of study. Specifically, female (total) participation rates are higher than men's in STEM fields, but the opposite happens in non-STEM disciplines. In light of these findings, it is not surprising that we did not find differences in (total) participation rates between men and women when pooling all fields of study. Namely, the offsetting gender gaps in STEM and non-STEM fields lead to gender-balanced overall participation rates in the LCF program (as in column 3 in Table 2).

3.3 Gender differences in GPA

The LCF fellowships are very prestigious and highly competitive. As a result, the vast majority of applicants have very high GPA and it has been shown that GPA is also one of the most important factors determining the probability of being awarded a fellowship (Farré and Ortega (2021)).

The goal of this section is to examine whether the gender differences in participation in the fellowship program can be explained by gender differences in GPA. For this to be the case, female students should be over-represented among the top students (by GPA) in STEM fields and the opposite should be true in non-STEM fields.

The administrative data on graduation records include each student's GPA (on a 0-10 scale with a passing grade of 5). We use these data to characterize each student's position in the GPA distribution corresponding to *his or her major and university*, which will account for any differences in grading standards across these dimensions. In order to provide a simple comparison of the resulting GPA distributions of male and female students, we first compute the density of male (female) students that fall into the percentiles of the GPA distribution for each major-university pair. Then, at each 5 percentage-point bracket, we compute the female-male gap. Obviously, if the GPA distributions for male and female students were identical (within major-university), the resulting line would be constant and equal to zero for all brackets.

The top panel in Figure 3 reports the resulting gender gap function when pooling all fields of study. The figure clearly shows that women are greatly under-represented at the bottom of the GPA distribution: the share of women with GPA below the 5th percentile is more than 1 percentage point lower than the corresponding share for males.

In fact, below the 25th percentile, the density of women is always lower. In contrast, females are over-represented between the 25th and 90th percentiles. However, women are again under-represented in the top brackets, particularly above the 95th percentile (by about 0.5 percentage points).[24]

Let us now turn to the GPA gender gaps by field of study depicted in the remaining panels of the figure, focusing on the higher GPA percentiles. The most striking observation is the large under-representation of women with grades above the 90th percentile in the Arts & Humanities. The figures also reveal that women are under-represented at the top of the GPA distribution in STEM fields. In contrast, female students are over-represented at the top of the GPA distribution in Health & Life sciences.[25] In Social sciences, women appear to be over-represented in the 90-95 percentile bracket but slightly under-represented in the 95-100 percentile bracket.

Summing up, with the exception of Health & Life sciences, female students are under-represented in the top GPA decile of their major-university. Accordingly, gender differences in GPA distributions are unable to explain the higher female participation rate in STEM and their lower participation rate in Health & Life sciences observed in the aggregate data (Table 2); factors other than cognitive skills (as measured by grades) must be playing an important role. In contrast, the marked low presence of females among top students in Arts & Humanities is likely to explain a great deal of the lower female participation in this field of study. We will come back to these questions in our analysis of the survey data.

3.4 Gender differences by program type and location

The LCF fellowship program funds both Ph.D. and Master's degrees. In addition, about 30% of the fellowships are given out through a sub-program that funds Ph.D. studies within Spain (in a province that differs from the one where the applicant attended college). This feature of the program allows us to investigate gender differences in

[24] Appendix Table A provides estimates of these gaps along with the corresponding standard errors. Column 3 shows that the share of women in the top half of the GPA distribution (by major-university) is 1.89 percentage-points higher than the share of men. Above the 75th percentile, women remain over-represented (by 1 percentage point). However, they are under-represented above the 90th percentile (column 5) by almost 0.5 percentage points. In fact, the share of female students is 0.39 percentage-points lower than men's in the top 2 percent of the pooled GPA distribution (column 7).

[25] Columns 8-12 in Appendix Table A estimate the gender gaps in GPA in the 95-100 percentile bracket, by field of study. On the basis of the statistically significant estimates, we find that females are under-represented in STEM (by 0.68 percentage points) and, particularly, in Arts & Humanities (by 1.4 percentage points).

preferences for geographical mobility and length of graduate studies.

The top panel in Table 5 reports the overall total participation rate (TPR), pooling all fields of study, separately by gender and type of program. As shown earlier, the TPR for both genders is 1.65%.[26] Comparing the two types of programs involving moving abroad, we observe that both men and women are similarly interested in Master's degrees (with a female-male TPR ratio of 1.01). However, females are much less interested in a Ph.D. abroad than men (by 15 percentage points). This indicates that, on the whole, females are less interested in Ph.D. programs than males.

Turning now to applications for domestic and foreign Ph.D. programs, the figures in the last two columns of Table 5 show that females are more interested than males in domestic programs, but less interested than them in foreign-based programs. Taken together, these two findings suggest that women are relatively less interested in Ph.D. programs than men and also less keen on moving abroad for their graduate studies.

The remaining panels report estimates by field of study. The estimates confirm the same pattern in all fields, with the exception of graduates in the field of Health & Life sciences. Namely, in STEM, Arts & Humanities and Social sciences, the female-male TPR ratio for Master's abroad is higher than for doctoral programs abroad, and the latter is lower than the female-male TPR ratio for doctoral programs in Spain. Thus, except in the field of Health & Life sciences where the ranking of TPR ratios is markedly different, female students are relatively less interested in doctoral programs and less geographically mobile than males.

4 The participation decision

The goal of this section is to investigate further the determinants of the decision to participate in the LCF program at the individual level. Importantly, one of the universities in our dataset (UB) agreed to merge their graduates' records with the LCF applications dataset. The resulting matched data allows us to estimate models of the decision to participate that partial out the roles of GPA, gender, age, socio-economic status (in the form of parental education), and field of study. In particular, these data allow us to estimate the *conditional* gender gap in participation net of differences in observable characteristics.

It is worth noting that the UB is the largest university in Catalonia in terms of

[26] According to our applications data, 52% of all applicants are interested in Master's degrees abroad, 17% in Ph.D. degrees abroad and 31% seek funds for Ph.D. programs in Spain.

enrollment, accounting for 47% of the enrollment records in our data. The average characteristics of the UB students are fairly similar to the other universities in our study, with the exception of the engineering school (UPC). However, the admissions cutoff for some of the UB's largest majors is lower than for the other broad-based universities in our study. As a result, the aggregate total participation rates for the UB are relatively low.[27] As we shall see soon, participation in the LCF fellowship program is heavily concentrated among the highest-GPA students. Variation in unobserved characteristics among the *top* students across the universities included in our study is likely to be much lower than across the *average* student in those universities. To the extent that this assumption is correct, the findings in this section will apply to the other universities as well.

Let us begin by providing some basic descriptive statistics (Table A.13). Roughly 66% of the UB graduates in our data are female and, on average, female graduates have a slightly higher GPA.[28] In addition, females are as likely as men to be between the 90th and 95th percentiles of their major GPA distribution, but 0.23 percentage points less likely to be in the top 5% than men. The Table also shows that female UB graduates are 0.8 years younger than men (thus, complete their majors faster) and less likely to have at least one college-educated parent.[29] In terms of fields of study, the composition of UB graduates departs somewhat from the values of the 4 universities pooled together. About 5% of all female UB graduates majored in STEM fields, which is approximately 11 percentage points lower than when combining the four universities (see bottom of Table 2). In contrast, female UB graduates are over-represented in Health & Life sciences and Social sciences (by about 5 percentage points in each of those fields).

Using our matched data, we can estimate linear probability and probit models on the sample of all UB graduates where the dependent variable is an indicator for whether each individual applied to the LCF program at any point after graduation. The models

[27] Appendix Table A.11 reports total participation rates (TPR) by university for cohorts 2012-2014. Pooling fields and genders, the TPR for the UB is 1.07%, substantially lower than for the other two broad-based universities (which is around 3%). But, importantly, the female-male ratios are fairly similar for the UB (0.90) and the other two broad universities (0.91 and 0.97).

[28] For comparison, the female shares among the other two broad-based universities in our data (UPF and UAB) over the same time period were 61% and 64%, respectively. In the three broad-based universities, female graduates had slightly higher GPA on average than male graduates. The gender gaps were very similar at the UB, UPF and UB: 0.21, 0.12 and 0.20 (on a 0-10 scale), respectively.

[29] The lower socio-economic status (SES) of female college graduates is not surprising given that the average SES of men and women is the same in the population at large, but male students account for only 1/3 of overall enrollment. Thus, male college students are more highly selected in terms of family background.

include controls for several relevant students' observable characteristics, such as cohort (which absorbs cohort shocks such as the state of the labor market upon graduation), position in the GPA distribution, age at graduation (which is another proxy for academic ability), family background (parents' educational attainment) and, of course, gender. As a result, the coefficient of the female dummy in our models will identify the gender gap in participation in the LCF program conditional on observable characteristics.

The top panel of Table 3 reports estimates of linear probability models that only include cohort dummies. Columns 1 and 2 are estimated on the full sample of students and suggest a slightly lower participation rate for female graduates (by about 0.58 percentage points when controlling linearly for GPA in column 2). Columns 3-6 restrict estimation to subsamples along the GPA distribution. Columns 3 and 4 make clear that there are practically no gender gaps below the 90th GPA percentile. Columns 5 and 6 restrict to the graduates with GPA in the top 10 and 5 percent of their major and university, respectively. Here, we find a large and significant female penalty (around 2 to 3 percentage points in terms of participation probability). The middle panel includes controls for age at graduation and parental education. The estimated coefficients for the female dummy are very similar to the top panel. Among top students, females graduates are much less likely to apply to the LCF program than male students with the same observable characteristics. The gender gap reaches 3.84 percentage points above the 95th GPA percentile, which amounts to more than half the mean participation rate for this group of graduates (6.12%).

The estimates also reveal that having college-educated parents plays an outsized role in shaping participation decisions: having two college-educated parents increases the probability of participation by almost 7 percentage points relative to graduates without college-educated parents, and by about 2.4 percentage points relative to graduates with a single college-educated parent. Interestingly, age at graduation is found to negatively affect the probability of participation. Generally speaking, older students at graduation may be academically weaker. However, this is less likely to be the case when we restrict the sample to students with GPA above the 95th percentile (column 6). In this case, the negative coefficient on age at graduation may indicate that older students are more interested in joining the labor market than on carrying out graduate studies at a distant location. Lastly, the bottom panel of the table shows that the estimated gender gaps are practically identical if we estimate a probit model instead of a linear probability model.

Let us now estimate field-specific conditional participation gaps. The estimates are collected in Table 4 and we restrict the estimation sample to students with GPA above the 90th percentile (except in column 1 where we consider graduates with GPA in the 75th to 90th percentile). Columns 1-3 strongly suggest that conditional participation rates are lower for female graduates in all fields of study, relative to male graduates with the same observable characteristics. Unfortunately, the precision of the point estimates is low due to small sample sizes when using only cohorts 2012-2014.

To increase the number of observations, columns 4-5 report estimates based on all cohorts (2009-2017), which increases the sample 6-fold in STEM and 9-fold in non-STEM fields. However, as discussed in detail earlier, increasing sample size (and gaining precision) comes at the cost of introducing some degree of censoring bias. At any rate, the point estimates based on the larger sample (columns 4-5) show that the estimated gender gap remains statistically insignificant in STEM whereas we can now reject the zero null hypothesis for non-STEM fields. Importantly, this change is primarily due to the large reduction in standard errors (since the point estimates remain similar to those obtained in columns 2-3), suggesting at most a small censoring bias. Disaggregating further the non-STEM fields, we find clear evidence of lower female participation (relative to observationally similar males) in Arts & Humanities, Social sciences and Health & Life sciences (although we cannot reject the zero null hypothesis for the latter). It is also helpful to examine Figure 4, which provides graphical illustration of these findings, making it clear that the gender gaps in participation appear only above the 90th GPA percentile in non-STEM fields.[30]

Summing up, our analysis of individual participation decisions using the UB-LCF matched data has led to clear evidence of lower participation rates among female graduates at the top of the GPA distribution, relative to male graduates with the same observable characteristics. Our estimates strongly suggest that this gender gap arises solely in non-STEM fields. Additionally, our results suggest that GPA, age at graduation and socio-economic background are all important determinants of the decision to participate in the LCF program. Moreover, the remaining conditional gender gaps in participation among the most brilliant students indicate that other factors are also relevant to understand students' plans after college graduation. In the next section, we analyze the information collected in our survey to identify other potential factors driving their post-graduation decisions.

[30]Once again, the magnitudes of the effects are very similar if we instead report marginal effects from probit models (columns 6 and 7).

5 Survey on post-graduation plans

In the previous sections we have documented gender gaps in participation in the LCF program among high-GPA students. It is important to recall that the main goal of the LCF program is to fund graduate studies in the world's best universities for academically brilliant students. As a result, gender gaps in application rates may reflect gender differences in preferences for graduate studies or in geographic mobility.

Administrative data is silent regarding the nature of the differences in program participation between male and female students. To explore students' interest on graduate studies, we conducted a large-scale online survey in the four universities participating in our study during January and February 2020. Specifically, this section presents summary statistics for the main variables of interest, explores gender gaps in the most relevant individual factors shaping the decision to study abroad, and estimate (conditional) gender gaps on interest over graduate studies and in the geographic location of each student's preferred graduate program.

5.1 Survey description

The population of interest were college students that had completed over half of the credit hours required for graduation. In total, we gathered approximately 4,000 complete questionnaires (as discussed in Section 2.3). The data show that 72% of students intended to enroll in graduate school and practically the same number (71%) reported intending to seek employment immediately after graduation. Presumably, these students had not yet firmed up their post-graduation plans and were considered multiple options. The survey contains a wealth of information to understand the factors shaping students' intended plans after graduation, including sociodemographic characteristics and information on attitudes toward graduate studies, professional careers and family.

Table 6 presents summary statistics, comparing men and women for the whole sample and the subsample with high grades (defined as above the 75th percentile on the basis of administrative data pooling all fields). The results in the table are largely in agreement with the administrative data discussed earlier. Female students are severely under-represented in STEM fields, but over-represented in non-STEM fields (and particularly in Social sciences). In addition, female students are slightly younger than men (by about 1/3 of a year) and have slightly lower socio-economic status, measured by parental education. In regards to academic achievement, the survey data show that women have slightly higher GPA, on average, than men, but are 4 percentage-points less likely to be in the top 10% by GPA.[31] When we restrict to students with high GPA, (above the 75th percentile), we observe that women are under-represented at the very top of the GPA distribution. In particular, among students with GPA above 75p, females are 14 percentage points less likely to be above the 90th percentile than male students. In terms of students' interest in graduate studies, about 69% of male students are interested in Master's studies and 12% are interested in doctoral programs. These figures rise to 72% and 23% when we restrict to the high-GPA sample (of male students). For this sample, females appear as interested as males in Master's degrees. However, the share of (high-GPA) females interested in Ph.D. programs is almost 10 percentage-points lower than the share of males.

5.2 Gender gaps in characteristics

There may be gender differences in characteristics that are important to determine career choices and, in particular, the decision to carry out graduate studies. As highlighted by a growing literature, these disparities may be endogenous responses to the *gender glass ceiling* in the labor market. For instance, female students may expect lower future earnings, which may discourage investing in post-graduate education (Bertrand and Hallock (2001), Bertrand (2013), Bertrand et al. (2019)). In addition, interest in graduate studies among women may be reduced if they have stronger preferences for children and family (Wiswall and Zafar (2021)) or if they are less geographically mobile than men (Le Barbanchon et al. (2019) and Petrongolo and Ronchi (2020)).

Besides providing a wide range of socio-demographic and academic information, our survey contains individual measures of earnings expectations (10 years after graduation), desired number of children (and age at which to have the first child), current family ties (in terms of hours providing care for siblings and older relatives and romantic relationships at the time of the survey). Next, we employ linear regression models to estimate whether the average male and female students in a given university and field of study differ along these dimensions.

The estimates are reported in Table 7. The top panel considers all students that answered the corresponding questions. We find statistically significant gender differences along five dimensions: female students are half a year younger, expect lower earnings (by 16 log points), would like to have children at an earlier age (by 1.3 years), spend

[31] Almost 20% of our sample reports grades above the 90th percentile (computed on the basis of the administrative data) because high-achieving students are more likely to respond to the survey.

more time in caregiving (by 0.6 hours per week) and are more likely to be in a romantic relationship (by 10 percentage points) than men in the same field of study and university.

A similar pattern arises when we restrict the analysis to students with GPA above the 75th percentile, although we now observe slightly lower GPA among females (by 0.12 points on a 0-10 scale) and a larger gender gap in caregiving hours (of 1.2 hours weekly). The bottom two panels consider separately STEM and non-STEM fields (for students with GPA above the 75th percentile). While the general pattern is similar in both areas of study, it is worth noting the much larger expected gender wage penalty and moderately larger higher prevalence of romantic relationships reported by female students in STEM relative to their male counterparts. In the non-STEM sample, what stands out is a larger gender gap in the preferred age at which to have the first child (i.e. on average females would like to have their first child 1.7 years earlier than males in the same field of study and university).

In sum, generally speaking, the data for high-GPA students reveal significant gender gaps along several dimensions. In particular, we found that women have slightly lower grades, substantially lower expected earnings, would like to have children at a younger age, and stronger involvement with their families and romantic partners than male students in the same field of study and university. As we shall see below, some of these characteristics help understand gender gaps in preferences over graduate studies. However, they do not shed much light on the gender differences in participation in the LCF program in STEM and non-STEM fields.

5.3 Interest in graduate studies

Let us begin by examining how many students are planning to pursue graduate studies after graduating from college. To answer this question, we consider the full sample of students and estimate a model where the dependent variable is an indicator taking a value of one if the student is interest in enrolling in a Master's or doctoral degree (at any location).

The results are reported in columns 1-3 of Table 8. The first column shows that 72% of students were considering pursuing graduate studies and we do not observe any significant gender differences. Column 2 restricts the sample to students with GPA above the 75th percentile. As expected, the share of students interested in graduate studies increases (by 3 percentage points) and, once again, we do not find evidence of gender differences. Column 3 includes a vector of characteristics in order to explore the factors that shape the decision to pursue graduate studies, but none of the estimates are statistically significant (with the exception of care hours, which appears to have a positive effect). The middle and bottom panels provide estimates for the sub-samples of STEM and non-STEM students, respectively. Two observations are worth noting. First, the share of students interested in graduate studies is much larger in non-STEM areas (95% versus 83%). Secondly, though not statistically significant, the estimates suggest that female students are more interested in graduate studies than observationally equivalent male students in STEM fields, but the opposite may be true in non-STEM fields.

All in all, these estimates underscore the widespread interest in graduate studies among college students of both genders. The estimates also hint at a relative higher interest among female students in STEM and a lower interest among female students in non-STEM areas (relative to observationally similar males). This pattern echoes the participation gender gaps estimated in Section 3 and Section 4.

Next, we restrict the analysis to high-GPA students (above the 75th percentile) who are *interested in graduate studies*. As shown in column 4, practically all of these students are potentially interested in enrolling in Master's programs (99% in STEM and 95% in non-STEM fields, respectively) and there is practically no difference between male and female students. Interestingly, when we include our vector of characteristics, we uncover a *conditional* gender gap. Namely, among students interested in graduate studies, females are 5 percentage-points *more* likely to consider pursuing Master's degrees after graduation, and the gender gap is more prominent in non-STEM fields. The estimates also uncover a negative (and statistically significant coefficient) for the romantic relationship indicator. Thus, a potential explanation for the increase in the gender gap relative to column 4 is the much larger prevalence of romantic relationships among female students (as seen in column 8 of Table 7).

Not surprisingly, interest in doctoral studies is much lower (columns 6-7): only 24% of students report having an interest in this type of program (rising to 27% in STEM).[32] However, it is striking that there is a very large gender gap: female students are 13 percentage points *less* interested in doctoral studies than male students with the same observable characteristics.[33] It is worth noting that the gender gap is present both in

[32]Though not shown in Table 8, the estimated coefficients for the field dummies included in our models reveal important cross-field differences in the relative preference for Master's versus Ph.D. degrees. Specifically, the estimates show that STEM students are 9 percentage-points more likely to report interest in a Master's degree (column 5) than similar students in Health & Life sciences, whereas students in Health & Life sciences are 5 percentage-points more likely to be interested in doctoral studies than STEM students.

[33]Interest in Master's and Ph.D. studies are not mutually exclusive categories. In fact, 21% of students

STEM and non-STEM fields, but it is substantially smaller in STEM. Additionally, GPA is found to be a significant determinant of the Master's versus Ph.D. choice: higher GPA students are more (less) likely to be interested in doctoral (Master's) studies.

In sum, the survey data clearly reveal that female college students are more interested in Master's programs and much less interested in doctoral studies than observationally similar male students. Generally speaking, this pattern arises both in STEM and non-STEM fields. Hence, STEM versus non-STEM disparities in gender differences in participation in the LCF fellowship program do not seem to arise from gender differences in interest in graduate studies.

5.4 Gender disparities in location graduate studies

Our survey asked students for the geographic location of their preferred graduate program. As shown in Table 9 (column 1), the majority of male and female students (61%) reported preferring a program located in the province where they currently reside.[34] About 11% of male students (and 15% of women) were considering programs in another province within Spain. However, while 28% of men reported planning to attend graduate school abroad, the corresponding value for women was 3.5 percentage points lower. Thus, among students planning to attend graduate school after graduating from college, women appear to be less willing to attend a foreign institution.

The survey also asked students about the location of their preferred graduate program in a *hypothetical* scenario where they did not face any *economic or family constraints*.[35] As shown in the second column of Table 9, students' 'unconstrained' choices would be dramatically different. The share of male students that would choose to study abroad would be 32 percentage-points higher. Among women, the increase in the corresponding share is 3 percentage-points higher than for men. In other words, the underlying preference for studying at a foreign institution seems to be the same for male and female students. However, economic or family restrictions constrain women more than men in terms of their geographical mobility. Given that the average wealth of the parents of male and female students is unlikely to differ much, this finding suggests that family considerations, broadly defined, may be responsible for the lower geographic mobility of

interested in graduate studies report being potentially interested both in Master's and doctoral degrees.

[34] The universities included in our study are located in the Barcelona metropolitan, which offers a wide variety of options for graduate studies.

[35] The exact wording of the question is the following: "In the absence of family and economic constraints, your preference would be to carry out graduate studies in (i) the current province of residence, (ii) in another province (within Spain), (iii) in another European country, or (iv) outside of Europe.

female students.

To investigate further the reasons for why the preferred location of the graduate program for females is less likely to be abroad than for male students, we estimate models that include the vector of characteristics described in Table 7. Our analysis is restricted to students reporting an interest in graduate studies, which is the vast majority. The dependent variable in the model is an indicator taking a value of one if the student's preferred graduate program is located outside of Spain.

The estimates are collected in Table 10. The first column refers to the location of the preferred graduate program in the hypothetical scenario were the student was unconstrained by family or economic restrictions, whereas all other columns refer to the actual (constrained) choice. Consistent with the findings in Table 9, we do not find significant differences between the locational preferences of male and female students in the unconstrained scenario. However, when considering actual choices, female students are 5 percentage-points less likely to want to study abroad than male students.[36]

From column 3 onward, we restrict the sample to students with strong academic credentials (i.e. GPA above the 75th percentile). As expected, the unconditional share of students interested in studying abroad is 6 percentage-points higher among high-GPA students (relative to 26% for the whole sample). Additionally, column 3 shows that the gender gap in actual locational preferences rises to 9 percentage-points.

Column 4 displays the estimates for our complete model of the intention to study abroad. Several characteristics included in the model are statistically significant and have the expected signs. In particular, the R-squared of the model is 0.17 and, thus, the model is much more successful at explaining individual variation in the outcome variable than when we tried to explain students' interest in graduate studies; the R-squared for that model was only 0.06 and only none of the individual characteristics was statistically significant (column 3 in Table 8). Most importantly, the model explains away the unconditional gender gap in locational preferences reported in column 3.

Let us examine further which individual characteristics are responsible for the unconditional gender gap in intentions to study abroad (documented in column 3). Several points are worth noting. Unsurprisingly, GPA and socio-economic status (SES) seem to have positive effects on the intention to study abroad, whereas age at the time of the survey has a small negative effect. As reported in Table 7, female students (in or

[36]Consistent with Table 9, the gender gap in column 2 (actual choice) is 3 percentage-points lower than in column 1 (unconstrained choice). The number of observations was lower in Table 9 because only students who responded to both questions could be included.

approaching senior year) are about half a year younger than male students. However, the difference was not statistically significant for students with high grades. However, gender differences in GPA and SES were found to be statistically significant. On the basis of our estimated slope coefficients, females' preference for studying abroad is reduced by 1.7 percentage points due to GPA gender differences (-0.12×0.14). This effect is partially offset by the slightly higher SES among females, which implies a 0.8 percentage-point increase in the intention to study abroad.

Interestingly, the coefficient for expected earnings is positive and statistically significant, and we found earlier that female students expect earnings that are 10 log points lower than what male students expect (among those with GPA above the 75th percentile). Combining this gender gap with our estimates implies that gender gaps in earnings expectations may lower females' intentions to study abroad by a full percentage point. Similarly, involvement in romantic relationships could also be playing an important role quantitatively for female students (but not for males), as suggested by the estimates in columns 5 and 6. Our estimates imply that gender gaps along this dimension reduce females' intention to study abroad by an additional 0.9 percentage points (-0.07×0.13). All in all, gender gaps in GPA, SES, expected earnings and romantic relationships account for a 2.75 percentage-point gap, which is almost one third of the 9 percentage-point unconditional gender gap in the intention to study abroad. Additionally, the estimates also suggest that females in STEM are much more likely to want to study abroad than observationally similar female students in other fields.

Naturally, the usual (and important) caveat to assign causal interpretations to findings based purely on cross-sectional analysis applies here as well. Our findings are largely correlational and we cannot rule out the existence of unobserved characteristics that are truly responsible for the seemingly significant coefficients of earnings expectations or romantic relationships.[37]

Last, columns 7 and 8 estimate our main model (column 4) separately on STEM and non-STEM students. The data show that STEM students are almost 20 percentage-points more interested in studying abroad than non-STEM students (46% versus 27%), probably reflecting the lower standardization of programs in non-STEM areas, which

[37] It is worth noting that we replicated the analysis in Table 8 but using as dependent variable the preference for studying abroad in the hypothetical unconstrained scenario. In theory, this question rules out a wide range of unobservable characteristics (though obviously not all). The most noteworthy finding is that the point estimates for the coefficient of the romantic relationships indicator grow larger and more statistically significant than in Table 8. This finding makes the causal interpretation a little more plausible. The results are available upon request.

includes majors such as law or education. The estimates further suggest that earnings expectations and romantic relationships are more important determinants (or correlate more strongly with the unobserved factors that are truly responsible for) the decision to study abroad for students in non-STEM fields. It is worth highlighting that the estimates in column 7 hint at a higher propensity to wish to study abroad for STEM females relative to observationally similar males in the same field of study. Together with the low share of females in STEM (29% in our administrative data), this finding suggests that the larger participation rate among females in STEM majors is related to a high degree of positive self-selection.

In sum, our analysis has shown that, on average, female students are more geographically constrained than men when considering (post-graduate) educational decisions. Furthermore, on the basis of gender gaps in the individual characteristics measured in our survey, we have been able to account for about one third of the unconditional gender gap in intention to study abroad. Besides gender gaps in GPA, SES and field of study, earnings expectations and romantic relationships could be playing a quantitatively significant role. These findings relate to Wiswall and Zafar (2021), which documented the role of earnings, marriage and fertility prospects on the educational choices of college students (regarding the choice of major).

6 Conclusions

Strong credentials, such as advanced degrees from renowned universities, help gain access top positions in the private and public sectors (including academia). At the same time, it is well established that women remain under-represented in high-status positions in the labor market and also in prestigious graduate programs across many fields of study.

Our paper analyzes the academic aspirations of male and female college graduates using administrative data on applications to a highly competitive fellowship program aimed at funding post-graduate education for candidates with excellent academic qualifications, combined with administrative graduation records for the largest universities in the region of Catalonia and a new large survey addressed to college students in these universities about their post-graduation plans.

Our analysis clearly shows that students with higher grades are much more likely to intend to pursue graduate studies abroad and to seek funding from competitive fellowship programs. The data also show that there are no significant gender differences in application rates to the fellowship program when pooling all fields of study. However, we document that female college graduates in STEM fields exhibit *higher* participation rates than male students in the same fields, while the opposite is true in non-STEM fields. Analysis based on individually matched data for the largest university in our sample shows that these gender differences are primarily found among high-GPA students (above the 90th percentile) even after accounting for socio-economic status and age at graduation.

Lastly our survey data shed some light on the interest in graduate studies and geographic preferences of male and female college students. We found that female students (both in STEM and non-STEM fields) are much less interested in doctoral studies than observationally similar males, and moderately more interested in master's programs. In addition, we also find that, among students interested in post-graduate studies, females (in non-STEM areas) tend to be less interested in studying abroad than males and that this gap is largely explained by their lower earnings expectations and higher involvement in romantic relationships.

References

Azmat, Ghazala and Katja Kaufmann, "Formation of College Plans: Expected Returns, Preferences, and Adjustment Process," Technical Report, Mimeo Sciences Po 2021.

_ and Rosa Ferrer, "Gender Gaps in Performance: Evidence from Young Lawyers," *Journal of Political Economy*, 2017, *125* (5), 1306–1355.

_ , Vicente Cuñat, and Emeric Henry, "Gender Promotion Gaps: Career Aspirations and Workplace Discrimination," 2020.

Bagues, Manuel F. and Berta Esteve-Volart, "Can Gender Parity Break the Glass Ceiling? Evidence from a Repeated Randomized Experiment," *Review of Economic Studies*, 2010, *77* (4), 1301–1328.

Barbanchon, Thomas Le, Roland Rathelot, and Alexandra Roulet, "Gender differences in job search: Trading off commute against wage," *Available at SSRN 3467750*, 2019.

Bayer, Amanda and Cecilia Elena Rouse, "Diversity in the Economics Profession: A New Attack on an Old Problem," *Journal of Economic Perspectives*, Fall 2016, *30* (4), 221–242.

Beneito, Pilar, Jose E. Bosca, Javier Ferri, and Manu Garcia, "Gender Imbalance across Subfields in Economics: When Does It Start?," *Journal of Human Capital*, 2021, *15* (3), 469–511.

Bertrand, Marianne, "Career, family, and the well-being of college-educated women," *American Economic Review*, 2013, *103* (3), 244–50.

_ and Kevin F Hallock, "The gender gap in top corporate jobs," *ILR Review*, 2001, *55* (1), 3–21.

_ , Sandra E Black, Sissel Jensen, and Adriana Lleras-Muney, "Breaking the Glass Ceiling? The Effect of Board Quotas on Female Labour Market Outcomes in Norway," *Review of Economic Studies*, 2019, *86* (1), 191–239.

Bielby, William T. and Denise D. Bielby, "I Will Follow Him: Family Ties, Gender-Role Beliefs, and Reluctance to Relocate for a Better Job," *American Journal of Sociology*, 1992, *97* (5), 1241–1267.

Black, Dan A, Amelia M Haviland, Seth G Sanders, and Lowell J Taylor, "Gender wage disparities among the highly educated," *Journal of human resources*, 2008, *43* (3), 630–659.

Blau, Francine D. and Lawrence M. Kahn, "The Gender Wage Gap: Extent, Trends, and Explanations," *Journal of Economic Literature*, September 2017, *55* (3), 789–865.

Boustan, Leah and Andrew Langan, "Variation in Women's Success across PhD Programs in Economics," *Journal of Economic Perspectives*, Winter 2019, *33* (1), 23–42.

Breda, Thomas and Son Thierry Ly, "Professors in Core Science Fields Are Not Always Biased against Women: Evidence from France," *American Economic Journal: Applied Economics*, October 2015, *7* (4), 53–75.

Bursztyn, Leonardo, Thomas Fujiwara, and Amanda Pallais, "'Acting Wife': Marriage Market Incentives and Labor Market Investments," *American Economic Review*, November 2017, *107* (11), 3288–3319.

Buser, Thomas, Muriel Niederle, and Hessel Oosterbeek, "Gender, competitiveness, and career choices," *The Quarterly Journal of Economics*, 2014, *129* (3), 1409–1447.

Cortes, Patricia, Jessica Pan, and Anna Sjogren, "Reaching for the Top: Gender Differences in the Labor Market and Household Outcomes of Top Talent in Sweden and the United States," Technical Report 2020.

Ellison, Glenn and Ashley Swanson, "The gender gap in secondary school mathematics at high achievement levels: Evidence from the American Mathematics Competi-tions," *Journal of Economic Perspectives*, 2010, *24* (2), 109–28.

Farré, Lídia and Francesc Ortega, "Selecting Talent: Gender Differences in Success in Competitive Selection Processes," *Journal of Human Resources*, 2021, pp. 0320–10767R1.

Fluchtmann, Jonas, Anita M. Glenny, Nikolaj Harmon, and Jonas Maibom, "The Gender Application Gap: Do men and women apply for the same jobs?," Technical Report 2020.

Fortin, Nicole M., Philip Oreopoulos, and Shelley Phipps, "Leaving Boys Behind: Gender Disparities in High Academic Achievement," *Journal of Human Resources*, 2015, *50* (3), 549–579.

Garcia-Montalvo, Jose, *Impacto de las becas 'La Caixa' de posgrado en el extranjero*, Obra Social 'la Caixa', 2014.

Goldin, Claudia, Lawrence F Katz, and Ilyana Kuziemko, "The homecoming of American college women: The reversal of the college gender gap," *Journal of Economic perspectives*, 2006, *20* (4), 133–156.

Hedges, LV and A Nowell, "Sex differences in mental test scores, variability, and numbers of high-scoring individuals," *Science*, 1995, *269* (5220), 41–45.

Hospido, Laura and Carlos Sanz, "Gender Gaps in the Evaluation of Research: Evidence from Submissions to Economics Conferences," IZA Discussion Papers 12494, Institute of Labor Economics (IZA) July 2019.

_ , Luc Laeven, and Ana Lamo, "The gender promotion gap: evidence from central banking," 2019.

Hsieh, Chang-Tai, Erik Hurst, Charles I Jones, and Peter J Klenow, "The allocation of talent and US economic growth," *Econometrica*, 2019, *87* (5), 1439–1474.

Jürges, Hendrik, "Gender ideology, division of housework, and the geographic mobility of families," *Review of Economics of the Household*, 2006, *4* (4), 299–323.

Keloharju, Matti, Samuli Knüpfer, and Joa cim Tåg, "What Prevents Women from Reaching the Top?," 2019.

Montalban, Jose and Almudena Sevilla, "The gender gap in student performance: the role of test-taking environment," Technical Report 2020.

Niederle, Muriel and Lise Vesterlund, "Do women shy away from competition? Do men compete too much?," *The Quarterly Journal of Economics*, 2007, *122* (3), 1067–1101.

Paredes, Valentina A., M. Daniele Paserman, and Francisco Pino, "Does Economics Make You Sexist?," NBER Working Papers 27070, National Bureau of Economic Research, Inc May 2020.

Petrongolo, Barbara and Maddalena Ronchi, "Gender gaps and the structure of local labor markets," *Labour Economics*, 2020, p. 101819.

Rouse, Cecilia and Claudia Goldin, "Orchestrating Impartiality: The Impact of 'Blind' Auditions on Female Musicians," *American Economic Review*, September 2000, *90* (4), 715–741.

Shauman, Kimberlee A. and Yu Xie, "Geographic mobility of scientists: Sex differences and family constraints," *Demography*, 1996, *33* (4), 455–468.

Wallen, Aaron S., Michael W. Morris, Beth A. Devine, and Jackson G. Lu, "Understanding the MBA Gender Gap: Women Respond to Gender Norms by Reducing Public Assertiveness but Not Private Effort," *Personality and Social Psychology Bulletin*, 2017, *43* (8), 1150–1170. PMID: 28903718.

Wiswall, Matthew and Basit Zafar, "Human capital investments and expectations about career and family," *Journal of Political Economy*, 2021, *129* (5), 1361–1424.

Table 1: Participation rates (PR) by cohort and calendar year

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Graduates	Applications	TPR (%)	PR (%)	PR (%)	PR (%)	PR (%)	PR (%)
Years		2014-18	2014-18	2014	2015	2016	2017	2018
Gender	both	both	both	both	both	both	both	both
Cohorts								
2009	13,024	26	0.20	0.09	0.01	0.05	0.02	0.02
2010	13,985	49	0.35	0.13	0.07	0.06	0.06	0.03
2011	12,428	77	0.62	0.21	0.13	0.08	0.10	0.10
2012	13,557	182	1.34	0.69	0.24	0.10	0.21	0.10
2013	**16,823**	**315**	**1.87**	**0.51**	**0.69**	**0.18**	**0.28**	**0.21**
2014	18,916	317	1.68	0.02	0.58	0.43	0.38	0.27
2015	18,942	271	1.43	0.02	0.02	0.51	0.54	0.35
2016	18,094	187	1.03	0.00	0.00	0.02	0.56	0.46
2017	17,109	106	0.62	0.00	0.00	0.01	0.01	0.61
All cohorts	142,878	1530	1.07	0.17	0.20	0.18	0.26	0.26
Cohorts 2012-14	**49,296**	**814**	**1.65**	**0.37**	**0.53**	**0.26**	**0.30**	**0.20**

Notes: The participation rate (PR) for graduating cohort c in year t is the number of applicants of that cohort in calendar year t (multiplied by 100) over the size of the cohort. The total participation rate in column 3 (TPR) sums applications by each cohort between calendar years 2014 and 2018. The 'ideal' cohort ($c = 2013$) contains students graduating in academic year 2013-2014. In the bottom panel, category *Cohorts 2012-2014* pools together cohorts $c = 2012, 2013, 2014$, which graduated in academic years 2012-2013, 2013-2014 and 2014-2015, and is our preferred estimate of the average TPR. Data pooled for the four universities (UB, UAB, UPC and UPF). For UAB we only use cohorts with complete enrollment data (2013-2014 onward).

Figure 1: Age-participation profile. Both genders combined

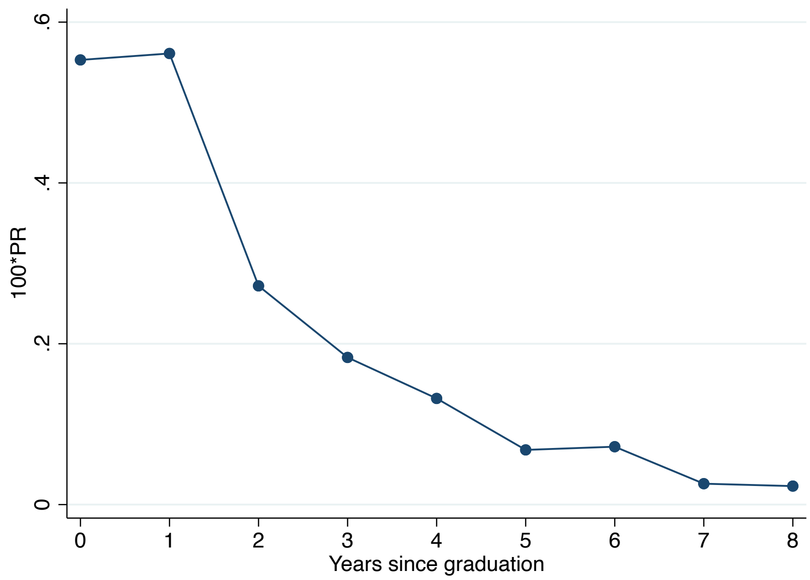
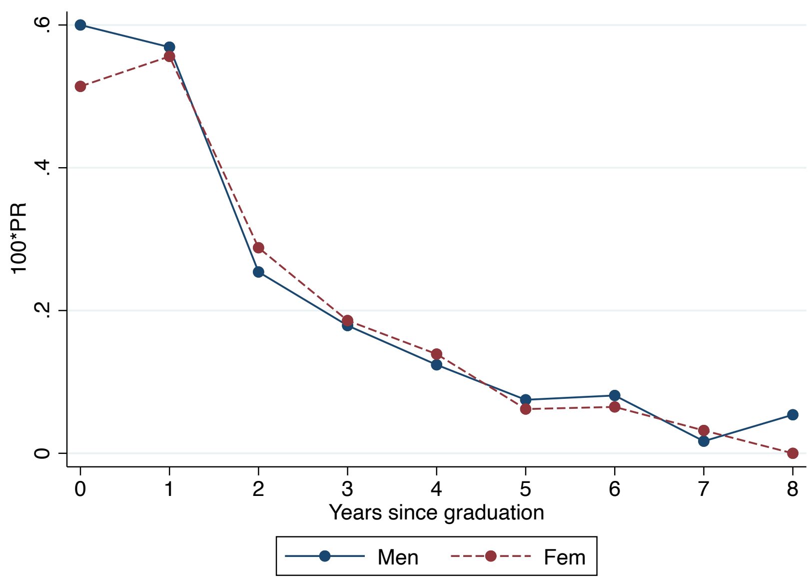
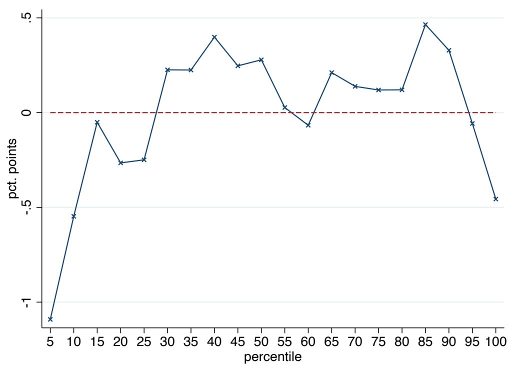


Figure 2: Age-participation profile, separately by gender

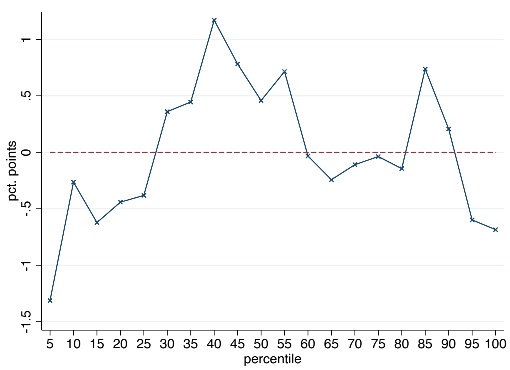


Notes: Average participation rate in the LCF fellowship program as a function of years since graduation. Both genders combined in the top figure and reported separately in the bottom figure. We report the number of participants per 100 college graduates, averaging across all cohorts. For a given cohort c in year$t$, the participation rate is defined as $PR_{c,t} = Applicants_{c,t}/Graduates_c$.

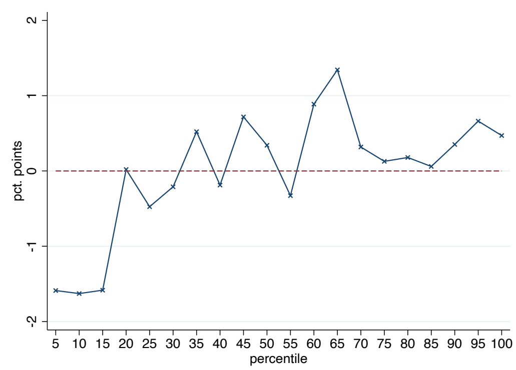
Figure 3: Female-Male gap in GPA distributions (in percentage points)



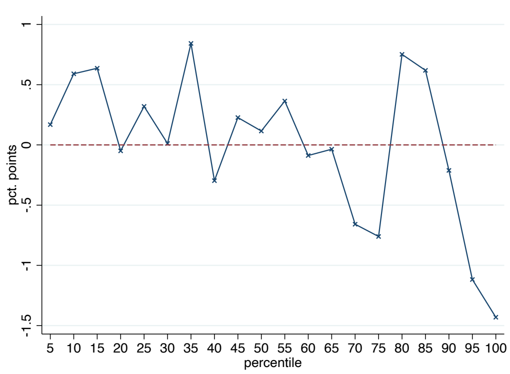
((a)) All fields



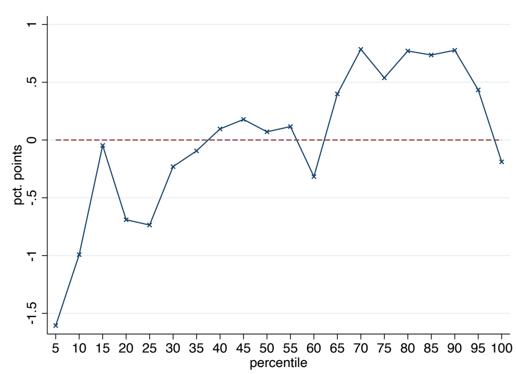
((b)) STEM



((c)) Health & Life Sciences



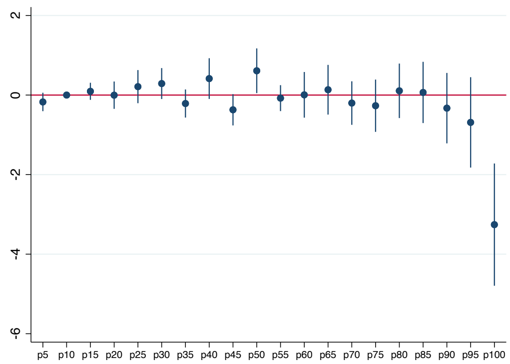
((d)) Arts & Humanities



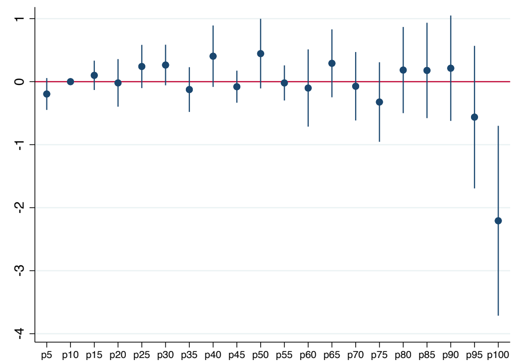
((e)) Social Sciences

Notes: Administrative data on graduation GPA for the 4 universities for graduation cohorts 2012-2014 (except for UAB for which we only use cohorts 2013-2014) for a total of 49,296 individual records. Each individual data point has been placed in the percentile bracket corresponding to the GPA distribution by major-university pair. The top panel pools all fields of study. The other panels restrict to the corresponding field of study.

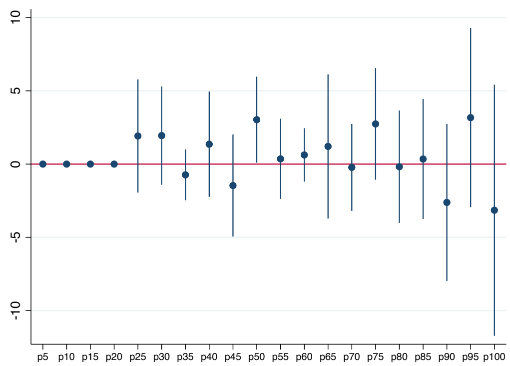
Figure 4: Female-Male gap in conditional participation. UB-LCF matched data



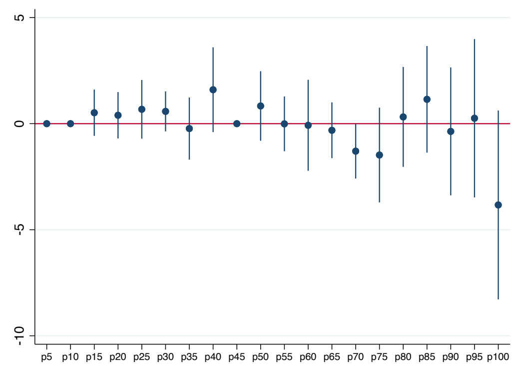
((a)) All fields



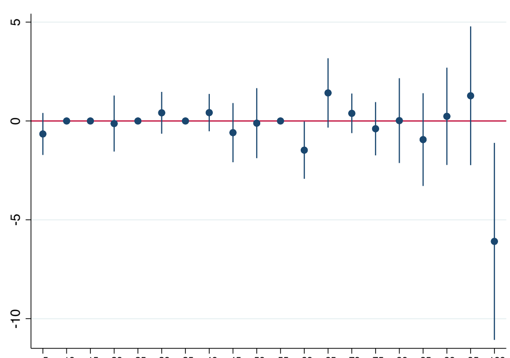
((b)) No STEM



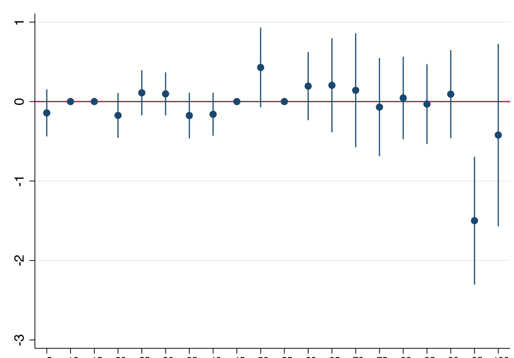
((c)) STEM



((d)) Health & Life Sciences



((e)) Arts & Humanities



((f)) Social Sciences

Notes: Linear probability model where the dependent variable is an indicator for applying to the LCF fellowship program (with a value of 100 if the student applied to the fellowship program). The right-hand side variables are age at graduation, GPA (if indicated), indicator for one parent with college education and indicator for both parents with college education. Model estimated on the matched UB-LCF dataset using graduation cohorts 2009-2017. Heteroskedasticity-robust standard errors.

Table 2: Total participation rates (TPR) by gender and field of study

Fields Variable	(1) All Grads	(2) All Applicants	(3) All TPR (%)	(4) STEM TPR (%)	(5) Health % Life TPR (%)	(6) Arts & Hum TPR (%)	(7) SocSci TPR (%)
All cohorts							
Both	142,878	1,530	1.07	1.18	1.94	1.73	0.54
Male	64,085	712	1.11	1.16	2.15	1.77	0.63
Fem	78,793	818	1.04	1.25	1.86	1.71	0.50
Fem/Male	1.23	1.15	0.94	1.08	0.86	0.96	0.79
Only 2013							
Both	16,823	315	1.87	1.86	3.82	2.93	1.03
Male	7,556	135	1.79	1.80	4.00	2.24	1.11
Fem	9,267	180	1.94	2.00	3.74	3.27	0.98
Fem/Male	1.23	1.33	1.09	1.11	0.94	1.46	0.88
Only 2013-2014							
Both	35,739	632	1.77	1.86	3.57	2.73	0.90
Male	16,300	290	1.78	1.79	3.88	2.63	1.04
Fem	19,439	342	1.76	2.02	3.45	2.78	0.83
Fem/Male	1.19	1.18	0.99	1.13	0.89	1.06	0.80
Only 2012-2014							
Both	49,296	814	1.65	1.67	3.45	2.65	0.86
Male	22,498	372	1.65	1.59	3.78	2.87	0.98
Fem	26,798	442	1.65	1.87	3.33	2.54	0.79
Fem/Male	1.19	1.19	1.00	1.18	0.88	0.88	0.80
Male % shares			100	50.3	7.9	8.4	33.3
Fem % shares			100	16.6	17.9	14.2	51.3

Notes: The total participation rate (TPR) in year t is the number of applicants over the period 2014-2018 (multiplied by 100) over the size of the corresponding graduating cohort. The top panel uses data for all cohorts in our dataset, starting from the cohort graduating in academic year 2009-2010 up to the cohort graduating in 2017-2018. Graduating cohort 2013 refers to students graduating in academic year 2013-2014. Graduating cohorts 2012-2014 refer to students graduating in academic years 2012-2013, 2013-2014 and 2014-2015. Male (Fem) % shares is the distribution of male (female) graduates (cohorts 2012-2014) by field. Data pooled for the four universities (UB, UAB, UPC and UPF). For UAB we only use cohorts with complete enrollment data (2013-2014 onward). Male and female shares across fields at the bottom of the table have been computed pooling cohorts 2012-2014.

Table 3: Individual participation models. Matched UB-LCF Data

Participate	(1)	(2)	(3)	(4)	(5)	(6)
Sample	All	All	$< 75p$	$75p - 90p$	$> 90p$	$> 95p$
Cohorts	2012-2014	2012-2014	2012-2014	2012-2014	2012-2014	2012-2014
LPM with Cohort dummies						
Female	-0.25	-0.58***	-0.04	-0.10	-2.20**	-3.11*
	[0.16]	[0.17]	[0.13]	[0.51]	[1.09]	[1.77]
LPM with Controls						
Female	-0.22	-0.55***	0.00	-0.18	-2.62**	-3.84**
	[0.17]	[0.17]	[0.13]	[0.51]	[1.09]	[1.78]
Age at grad.	-0.08***	-0.08***	-0.04***	-0.09***	-0.25***	-0.41***
	[0.01]	[0.01]	[0.01]	[0.02]	[0.04]	[0.07]
College parent 2	1.90***	1.82***	1.13***	1.59**	6.40***	6.95***
	[0.27]	[0.26]	[0.24]	[0.69]	[1.41]	[2.13]
College parent 1	0.56***	0.58***	0.11	1.48**	2.73**	4.58**
	[0.19]	[0.18]	[0.13]	[0.67]	[1.25]	[2.24]
mfx Probit with Controls						
Female	-0.31	-0.47	-0.05	-0.31	-2.34**	-3.37**
	[0.15]	[0.16]	[0.13]	[0.49]	[0.96]	[1.55]
Observations	18,195	18,195	13,638	2,756	1,801	915
Mean dep. var. %	1.04	1.04	0.5	1.52	4.39	6.12
GPA as control	No	Yes	No	No	No	No

Notes: The dependent variable is a dichotomous variable taking value of 0 or 100. The latter indicates that the individual applied to the LCF Fellowship program (in any year). In top and medium panels, the estimates are based on linear probability models and estimated on a sample containing only cohorts 2012-2014 (except for roughly 5,000 observations lacking SES) and pooling all fields of study. The GPA percentiles have been computed based on the administrative data for each major. Columns 3-6 restrict the sample to the graduates in the corresponding GPA percentile brackets. The middle panel includes cohort dummies, age at graduation, GPA (in column 2 only), and indicators for having exactly one parent with college education or two parents with college education. The bottom panel reports marginal effects (at the mean) based on probit models with the full set of cohort dummies and control variables (not shown for lack of space). Heteroskedasticity-robust standard errors.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 4: Individual participation by field of study. Matched UB-LCF Data

Participate	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Cohorts	2012-14	2012-14	2012-14	2009-17	2009-17	2009-17	2009-17
Model	LPM	LPM	LPM	LPM	LPM	Probit mfx	Probit mfx
GPA restriction	75p-90p	$> 90p$	$> 95p$	$> 90p$	$> 95p$	$> 90p$	$> 95p$
STEM							
Female	0.63	-0.71	-0.66	-0.41	-3.15	-0.55	-5.56
	[3.26]	[7.28]	[11.75]	[2.63]	[4.07]	[2.95]	[4.86]
Observations	209	139	73	479	240	479	240
Share of females %	41.15	30.94	35.62	35.28	36.25	35.28	36.25
Mean Dep. Var.	5.26	15.83	21.92	8.98	12.50	8.98	12.50
Non-STEM							
Female	0.18	-1.45	-2.54	-1.45***	-2.21***	-1.45**	2.07**
	[0.46]	[1.02]	[1.69]	[0.53]	[0.83]	[0.46]	[0.70]
Observations	2,547	1,139	571	5159	2624	5159	2624
Share of females %	70.48	68.53	67.81	68.79	68.06	68.79	68.06
Mean Dep. Var.	1.22	3.43	4.75	2.66	3.51	2.66	3.51
Health & Life Sciences							
Female	-0.64	-2.53	-6.94	-1.97	-3.83	-1.36	-3.04
	[1.77]	[3.30]	[5.56]	[1.62]	[2.50]	[1.37]	[1.98]
Observations	550	390	201	1224	629	1224	629
Share of females %	77.45	76.67	75.12	75.57	74.40	75.57	74.40
Mean Dep. Var.	2.73	7.44	9.95	5.31	6.68	5.31	6.68
Arts&Humanities							
Female	0.19	-4.04	-8.62*	-2.79*	-6.09**	-2.73*	-5.90**
	[1.72]	[3.12]	[5.02]	[1.66]	[2.71]	[1.49]	[2.34]
Observations	419	292	142	880	439	880	439
Share of females %	67.06	57.53	59.86	59.66	58.31	59.66	58.31
Mean Dep. Var.	3.10	6.16	8.45	5.23	7.29	5.23	7.29
Social sciences							
Female	0.27*	-0.5	0.43	-0.95**	-0.42	-0.88**	-0.49
	[0.16]	[0.75]	[1.13]	[0.43]	[0.62]	[0.36]	[0.69]
Observations		980	499	3055	1556	3055	1556
Share of females %	68.95	68.57	67.13	68.71	68.25	68.71	68.25
Mean Dep. Var.	0.19	1.02	1.60	0.85	1.16	0.85	1.16

Notes: The dependent variable is a dichotomous variable (with values 0 or 100) indicating if the individual applied to the LCF program (in any year). Columns 4-7 based on sample including cohorts 2009 through 2017. Linear probability models except in the last two columns, where we report the marginal effects resulting from probit models (with the same controls) evaluated at the sample means. The GPA percentiles have been computed based on the administrative data for each major. Control variables included in all models: age at graduation, an indicator for having exactly one parent with college education and an indicator for having two parents with college education. Heteroskedasticity-robust standard errors. *** p<0.01, ** p<0.05, * p<0.1.

Table 5: LCF participation rates by type of degree and location of graduate program

TPR	All sub-programs	Master Abroad	PhD Abroad	PhD Spain
All fields				
Male	1.65	0.86	0.31	0.49
Fem	1.65	0.87	0.26	0.52
FM ratio	1.00	1.01	0.85	1.06
STEM				
Male	1.59	0.81	0.30	0.48
Fem	1.87	0.85	0.28	0.74
FM ratio	1.18	1.05	0.93	1.55
Health & Life				
Male	3.78	1.11	0.37	2.29
Fem	3.33	1.21	0.63	1.48
FM ratio	0.88	1.09	1.69	0.65
Arts & Hum				
Male	2.87	1.72	0.68	0.47
Fem	2.54	1.62	0.36	0.56
FM ratio	0.88	0.94	0.53	1.20
SocSci				
Male	0.98	0.66	0.20	0.12
Fem	0.79	0.56	0.11	0.12
FM ratio	0.80	0.84	0.54	1.01

Notes: The total participation rate (TPR) in year t is the number of applicants over the period 2014-2018 (multiplied by 100) over the size of the corresponding graduating cohort. We only used the (mostly) uncensored cohorts 2012-2014, that is, students graduating in academic years 2012-2013, 2013-2014 and 2014-2015. Data pooled for the four universities (UB, UAB, UPC and UPF). For UAB we only use cohorts with complete enrollment data (2013-2014 onward).

Table 6: Survey descriptive statistics. All universities combined

	(1)	(2)	(3)	(4)
Sample	All	All	GPA$>75p$	GPA$>75p$
Gender	Men	Fem - Male	Men	Fem - Male
STEM	0.585	-0.361***	0.485	-0.338***
Health & Life	0.092	0.142***	0.135	0.093***
Arts & Hum	0.064	0.108***	0.097	0.155***
Soc. Sci	0.236	0.123***	0.204	0.131***
Age	23.815	-0.324***	23.765	-0.235
High SES	0.659	-0.035***	0.731	-0.011
GPA	7.057	0.188***	8.464	-0.121***
GPA$>75p$	0.283	0.008	1.000	0
GPA$>90p$	0.192	-0.041***	0.715	-0.145***
Plans Master	0.686	0.011	0.719	-0.013
Plans PhD	0.122	-0.037***	0.234	-0.099***

Notes: The table is based on the 4,848 individual questionnaires with non-missing data for gender, field of study and GPA. Columns 1 and 2 report statistics for students with any GPA, whereas columns 3 and 4 condition on the 1,093 students (22.5% of the sample) with GPA above the average 75th percentile across all fields of study in the administrative data. Columns 1 and 3 report means (for the corresponding samples) for men only. Columns 2 and 4 report the female-male difference in means for each variable. Thus, a negative value indicates a lower average value for females. The stars correspond to the test of no-difference in means. The first four variables are indicators of field of study. Hence, the mean corresponds to the share of students (of the corresponding gender) in that field. *High SES* is an indicator taking a value of one if at least one parent has a college degree. *GPA* is computed on a 0-10 scale, with 5 being the passing grade. *GPA¿75p* and *GPA¿90p* are indicators of GPA above the 75th and 90th percentiles, respectively. The *Plans* indicator variables (referring to Master's degree and PhD) are not mutually exclusive. Heteroskedasticity-robust standard errors. *** p<0.01, ** p<0.05, * p<0.1

Table 7: Survey gender gaps in characteristics

Dep. Var.	(1) Age	(2) GPA	(3) High SES	(4) ln *exp.earn.*	(5) Des. nChild	(6) Age child1	(7) Care h.	(8) Rom. rel.
All Fields								
Female	-0.53***	-0.02	0.01	-0.16***	0.07	-1.34***	0.60**	0.10***
	[0.15]	[0.03]	[0.02]	[0.02]	[0.04]	[0.16]	[0.28]	[0.02]
Obs.	3899	3,848	3,899	3,869	2,802	1,786	3,694	3,880
Mean DV	23.59	7.19	0.52	10.02	1.87	30.36	5.87	0.46
All Fields GPA>75p								
Female	-0.44	-0.12***	0.06*	-0.10***	0.07	-1.32***	1.17**	0.13***
	[0.34]	[0.03]	[0.03]	[0.04]	[0.09]	[0.39]	[0.56]	[0.04]
Obs.	953	902	953	947	671	404	907	948
STEM GPA>75p								
Female	0.17	-0.13*	-0.05	-0.24***	-0.01	-0.51	2.02*	0.19***
	[0.58]	[0.07]	[0.07]	[0.08]	[0.21]	[0.78]	[1.09]	[0.07]
Obs.	247	196	247	244	156	100	234	244
NoSTEM GPA>75p								
Female	-0.59	-0.11***	0.10**	-0.06	0.11	-1.67***	0.88	0.11**
	[0.42]	[0.03]	[0.04]	[0.04]	[0.10]	[0.47]	[0.66]	[0.04]
Obs.	706	706	706	703	515	304	673	704
FE field,uni	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: University and field of study fixed-effects included in all models. Except for first panel, all others include only students with GPA above 75p. Panel 4 includes field fixed-effects. The dependent variables are Age, GPA (0-10 scale), indicator for having 2 parents with college degrees, log of expected earnings (annual), desired number of children (*Des.nChild*), age at which would like to have the first child, number of hours per week taking care of relatives (children, grandparents, etc.), and an indicator for being in a romantic relationship. Heteroskedasticity-robust standard errors. *** p<0.01, ** p<0.05, * p<0.1

Table 8: Survey interest in graduate studies

Dep.var.	(1) Postgrad	(2) Postgrad	(3) Postgrad	(4) Master	(5) Master	(6) PhD	(7) PhD
All Fields							
Female	0.01	-0.02	-0.03	0.01	0.05**	-0.14***	-0.14***
	[0.01]	[0.03]	[0.04]	[0.02]	[0.02]	[0.03]	[0.05]
GPA			0.00		-0.03		0.13**
			[0.04]		[0.03]		[0.05]
High SES			0.02		-0.02		0.02
			[0.04]		[0.02]		[0.04]
ln $exp.earn.$			-0.02		-0.01		-0.04
			[0.03]		[0.03]		[0.05]
Care hours			0.01***		-0.00		0.00
			[0.00]		[0.00]		[0.00]
Rom. Rel.			0.04		-0.04**		-0.03
			[0.04]		[0.02]		[0.04]
Des. nChild			-0.00		0.01		0.01
			[0.02]		[0.01]		[0.02]
Observations	3,899	953	618	718	457	718	457
R-squared	0.00	0.00	0.06	0.00	0.05	0.03	0.07
Mean DV	0.72	0.75	0.75	0.96	0.96	0.24	0.24
STEM							
Female	0.05**	0.07	0.03	0.01	0.01	-0.10*	-0.13
	[0.02]	[0.05]	[0.08]	[0.01]	[0.03]	[0.06]	[0.10]
Observations	1,400	247	126	206	108	206	108
Mean DV	0.74	0.83	0.83	0.99	0.99	0.27	0.27
NoSTEM							
Female	0.01	-0.01	-0.04	0.03	0.06**	-0.17***	-0.15***
	[0.02]	[0.04]	[0.05]	[0.02]	[0.03]	[0.04]	[0.05]
Observations	2,499	706	492	512	349	512	349
Mean DV	0.71	0.72	0.95	0.95	0.95	0.23	0.23
GPA	All	$> 75p$	$> 75p$	$> 75p$	$> 75p$	$> 75p$	$> 75p$
Sample	All	All	All	Postgrad	Postgrad	Postgrad	Postgrad
FE field,uni	No	No	Yes	No	Yes	No	Yes

Notes: Columns 4-7 restrict sample to students interest in graduate studies. Considering to pursue a Master's degree (columns 4-5) and considering to pursue a Ph.D. (columns 6 and 7) are not mutually exclusive. When indicated, the models include indicators for university and broad field of study (STEM, Health & Life, Arts & Hum. and Social sciences). Controls defined in Table 7. Heteroskedasticity-robust standard errors. *** p<0.01, ** p<0.05, * p<0.1

Table 9: Survey. Preferred post-graduate program abroad

	(1) Actual pref. Percent	(2) Unconstrained pref. Percent	(3) Actual - Uncons. Percent
Males			
My province	61.3	32.8	-28.6
Other province	11.0	7.2	-3.8
Abroad	27.7	60.1	32.4
Europe	21.2	35.6	14.4
Outside Europe	6.5	24.5	18.0
sum	100	100	0
obs.	1,109	1,029	
Females			
My province	60.7	31.6	-29.1
Other province	15.1	8.7	-6.4
Abroad	24.2	59.7	35.6
Europe	19.4	37.8	18.4
Outside Europe	4.8	22.0	17.2
sum	100	100	0
obs.	1,554	1,458	

Notes: Column 1 reports answers to a question about students' actual preferences over the location of their preferred graduate program. Column 2 refers to a hypothetical question about students' locational preferences in a scenario where they did not face economic or family constraints. The sample is only those students who intend to pursue graduate studies. Heteroskedasticity-robust standard errors. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 10: Survey. Interest in graduate studies abroad

Abroad	(1) Uncons.	(2) Actual	(3) Actual	(4) Actual	(5) Actual	(6) Actual	(7) Actual	(8) Actual
Female	-0.02	-0.05***	-0.09**	0.01			0.10	-0.02
	[0.02]	[0.02]	[0.04]	[0.05]			[0.14]	[0.06]
Age				-0.01*	-0.01	-0.02	0.00	-0.01***
				[0.01]	[0.01]	[0.01]	[0.02]	[0.01]
GPA				0.14**	0.15*	0.14*	0.20*	0.12*
				[0.06]	[0.08]	[0.08]	[0.11]	[0.07]
High SES				0.14***	0.21***	0.11	0.24*	0.12**
				[0.05]	[0.06]	[0.09]	[0.13]	[0.05]
ln $exp.earn.$				0.11***	0.07	0.10**	0.02	0.12***
				[0.04]	[0.06]	[0.04]	[0.08]	[0.04]
Care hours				0.00	0.00	0.00	-0.01	0.00
				[0.00]	[0.00]	[0.01]	[0.01]	[0.00]
Rom. Rel.				-0.07	-0.13**	-0.01	0.08	-0.11**
				[0.05]	[0.06]	[0.08]	[0.13]	[0.05]
Des. nChild				0.01	0.04	-0.04	-0.03	0.02
				[0.02]	[0.03]	[0.03]	[0.05]	[0.03]
STEM				0.26***	0.23*	0.13		
				[0.09]	[0.14]	[0.14]		
HealthLife				-0.18***	-0.09	-0.45***		-0.18***
				[0.06]	[0.07]	[0.11]		[0.06]
ArtsHum				-0.01	0.08	-0.23*		0.01
				[0.06]	[0.08]	[0.13]		[0.07]
Obs.	2,098	2,215	544	354	219	135	79	275
R-squared	0.00	0.00	0.01	0.17	0.18	0.28	0.22	0.16
Mean DV	0.61	0.26	0.32	0.32	0.28	0.37	0.46	0.27
GPA	All	*All*	$> 75p$	$> 75p$	$> 75p$	$> 75p$	$> 75p$	$> 75p$
Gender	All	All	All	All	Fem	Male	All	All
Field	All	All	All	All	All	All	STEM	NoSTEM

Notes: Models estimated on the sample of students interested in pursuing graduate programs. The dependent variable is a dummy variable taking a value of one if the preferred graduate program is located abroad. Column 1 refers to the unconstrained hypothetical preference and the rest of columns to the actual preference. *Des.nChild* is the desired number of children. Columns 4-8 include university fixed-effects (not shown for lack of space). Heteroskedasticity-robust standard errors. *** p<0.01, ** p<0.05, * p<0.1

Appendix

A Tables and Figures

Table A.11: Total Participation Rates by University. Cohorts 2012-2014

Gender	Uni	Counts Applications	Counts Grads	TPR All Fields	TPR STEM	TPR Health	TPR ArtsHum	TPR Soc. Sci.
Both	All	886	49,296	1.80	1.70	3.97	2.87	0.95
Male	All	400	22,498	1.78	1.62	4.46	3.03	1.12
Fem	All	486	26,798	1.81	1.92	3.80	2.80	0.87
Fem/Male	All	1.22	1.19	1.02	1.18	0.85	0.92	0.77
Both	UB	249	23,217	1.07	2.75	2.30	1.41	0.33
Male	UB	93	8,087	1.15	2.43	2.86	1.53	0.33
Fem.	UB	156	15,130	1.03	3.22	2.11	1.35	0.34
Ratio F/M	UB	1.68	1.87	0.90	1.32	0.74	0.88	1.03
Both	Uni2	203	6,984	2.91	6.12	6.03	2.79	1.26
Male	Uni2	74	2,400	3.08	4.75	5.61	2.68	1.58
Fem.	Uni2	129	4,584	2.81	8.94	6.22	2.84	1.12
Ratio F/M	Uni2	1.71	1.91	0.91	1.88	1.11	1.06	0.71
Both	Uni3	173	5,517	3.14	5.71	10.38	4.79	2.17
Male	Uni3	68	2,132	3.19	4.61	8.57	5.52	2.47
Fem.	Uni3	105	3,385	3.10	11.76	11.05	4.60	1.97
Ratio F/M	Uni3	1.54	1.59	0.97	2.55	1.29	0.83	0.80
Both	Uni4	187	13,389	1.40	1.40			
Male	Uni4	134	9,746	1.37	1.37			
Fem.	Uni4	53	3,643	1.45	1.45			
Ratio F/M	Uni4	0.40	0.37	1.06	1.06			

Notes: The total participation rate (TPR) in year t is the number of applicants over the period 2014-2018 (multiplied by 100) over the size of the corresponding graduating cohort. Graduating cohort 2013 refers to students graduating in academic year 2013-2014. Graduating cohorts 2012-2014 refer to students graduating in academic years 2012-2013, 2013-2014 and 2014-2015. *All* refers to the four universities (UB, UAB, UPC and UPF) pooled together. For UAB we only use cohorts with complete enrollment data (2013-2014 onward). Except for the UB, the other universities are renamed to preserve confidentiality. The university specialized in engineering has a few majors in Social Sciences but they are very small in terms of enrollment. In fact, our data contain only one applicant to the fellowship program from these majors, which is insufficient to estimate application rates with any degree of confidence.

Table A.12: Comparison GPA distributions. Admin records all universities.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Dep.Var.	GPA	GPA	$> p50$	$> p75$	$> p90$	$> p95$	$> p98$	$> p95$	$> p95$	$> p95$	$> p95$	$> p95$
Fem	0.25***	0.10***	1.89***	1.01**	-0.48*	-0.44**	-0.39***	-0.68*	-0.33	0.45	-1.42**	-0.26
	[0.01]	[0.01]	[0.49]	[0.42]	[0.29]	[0.21]	[0.14]	[0.38]	[0.25]	[0.63]	[0.63]	[0.31]
Obs.	49296	49296	49296	49296	49296	49296	49296	16136	33160	6063	5815	21282
Mean DV	7.10	7.10	0.50	0.25	0.10	0.05	0.02	0.05	0.05	0.05	0.05	0.05
FE cohort	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FE uni	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FE field	No	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	No	No	No
Field	All	All	All	All	All	All	All	STEM	No STEM	Life	Hum	Soc
Uni	All	All	All	All	All	All	All	All	All	All	All	All

Notes: This table is based on graduation records for all four universities. We restrict to graduating cohorts 2012-2014 (except for the UAB for which we only use cohorts 2013-2014 due to missing data). The dependent variable in columns 1 and 2 is the GPA of each student (on a 0-10 scale, with passing grades above 5). The dependent variables in all other columns are based on each student's position in the GPA distribution of his/her major and university. Specifically, indicators (dummy variables) for belonging to the top (50, 25, 10, 5 or 2) percent of the GPA distribution. To reduce the number of decimals in the estimates, the indicators have been rescaled so that when a student satisfies the condition, the indicator takes a value of 100 (rather than the usual 1) and zero otherwise. Heteroskedasticity-robust standard errors.

*** p<0.01, ** p<0.05, * p<0.1

Table A.13: Summary statistics UB-LCF data

	All	Male	Fem-Male
Female	66.39		
GPA	7.24	7.10	0.21***
90p <GPA< 95p	5.19	5.17	0.03
GPA > 95p	5.21	5.37	-0.23
Age at grad.	24.28	24.81	-0.80***
One College parent	20.85	21.80	-1.43***
Two College parent	19.90	22.68	-4.20***
STEM	8.27	14.50	-9.38***
Health & Life	21.31	15.64	8.54***
Social sci.	55.25	54.09	1.75***
Arts & Hum.	15.00	15.60	-0.90***
Observations	63,701		

Notes: Matched UB-LCF dataset containing graduation cohorts 2009-2018. The sample includes only students with information on parental education (93% of the data). GPA is on a 0-10 scale (with passing grade 5.0). Indicator variables with GPA percentiles and parental background variables take a value of 100 if the student satisfies the condition and zero otherwise, which allows the coefficients to be interpreted as percentage points. GPA percentiles are computed within each student's major, pooling men and women. There are 73 majors at the UB.