QUEEN MARY UNIVERSITY OF LONDON

Gender Inequalities and Scarring Effects in School to Work Transitions

by

Silvia Granato

A Thesis submitted for the degree of
Doctor of Philosophy (Ph.D.) in Economics

in the

School of Economics and Finance
Queen Mary University of London

June 2018

Declaration of Authorship

I, Silvia Granato, confirm that the research included within this thesis is my own work or, where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

Abstract

This thesis investigates issues related to gender inequalities and scarring effects in school to work transitions.

The first chapter analyses the gender earnings gap among Italian college graduates at the beginning of their careers. Thanks to the richness of the dataset used I am able to control for a large set of variables related to individuals' educational and family background, as well as personality traits. The main finding is that the content of the college degree course is the most significant variable in explaining the earnings gender differentials of young workers. In particular I show that female sorting in college majors characterised by a low maths content explains between 13 and 16% of the earnings gender gap.

Motivated by this result, in Chapter 2 I investigate the determinants of gender gaps in STEM (science, technology, engineering and mathematics) graduation rates, with an emphasis on family, cultural and school influences. I show that half of the gap is attributed to the gender difference in maths and science content of the high school curriculum. The results indicate that in Italy the issue of the gender gap in STEM graduation has its roots in a gendered choice that originates many years before.

The final chapter analyses the extent to which the mismatch of demand and supply of skills that young workers face when they enter the labour market upon completing education affects their careers. Regression results show that there is a long lasting negative effect of these initial conditions on labour market outcomes. The evidence is suggestive of a 'trickle down unemployment' phenomenon, namely that high-skill workers try to escape strong competition from their high-skill peers by taking jobs for which a lower level of education is required, moving down the occupational ladder.

Acknowledgements

I am very grateful to my supervisor Barbara Petrongolo for the invaluable support and the enormous amount of time she dedicated to me in these years.

During my time at Queen Mary University of London I benefited from many useful discussions with the members of the faculty. In particular, I want to thank Erich Battistin for his precious and very valuable advice. I want to thank also my PhD colleagues, especially Valerio for giving me confidence and support, and Giulia with whom I shared the challenges of this journey; discussing research with her has always been very helpful, and without her I would probably not have made it through here.

A special thanks goes to the research institution AlmaLaurea, which made the realisation of most of this thesis possible. I am extremely grateful for everything they did for me in the last two years, from hosting me in their institution to giving me all the technical support that I needed. In particular I thank Davide Cristofori for his essential help during these years, from the first moment I joined AlmaLaurea to the last day before the submission of this thesis.

Of course the greatest gratitude goes to my amazing family, my parents Anna and Gino, and Chiara, *sorella indivisibile.* What they did for me in these years is much more than extraordinary, and I feel incredibly lucky to know what true unconditional love means thanks to them.

Last but not least, many thanks go to my "London family", Mirela, Maya and Serena, with whom I shared some of my best moments of these years.

Contents

List of Figures

List of Tables

Chapter 1

The Gender Earnings Gap Among Young Italian Graduates

1.1 Introduction

Last century has been characterised by a striking increase in women's participation to the economy. However, gender differentials in the labour market are still significant and persistent: important gaps remain in earnings and hours worked, and women are under-represented in high-status/high-income occupations. In 2016 the gender wage gap across OECD countries was 14.07% and the labour force participation of women was only 51.9% as opposed to 69% for men.[1]

Several studies have tried to investigate the determinants of the remaining gaps. The explanations explored are related to: (i) differences in maths ability and human capital; (ii) children and home production (workforce interruptions for motherhood, unequal division of housework and care responsibilities); (iii) occupational segregation; (iv) pay discrimination; (v) differences in preferences and psychological traits.

In this study I investigate the gender earnings gap among recent Italian college graduates at the beginning of their careers, with an emphasis on explanations related to the sorting of females and males into different fields of study and to gender differences in psychological traits. In Italy the gender pay difference in 2014 was approximately 5.6% of males median annual earnings, and women's employment

[1] Source: OECD (2015a). The gender wage gap is defined as the difference between median annual earnings of men and women relative to median annual earnings of men.

rate was 17.1 percentage points lower than the one of men. Figure 1.1 plots the employment and pay gap of the EU-14 countries in 2014 obtained from the OECD (2015a): the relatively lower wage gap in Italy is related to the sizeable selection into employment of women, as indicated by the relatively higher employment gap.

Many studies from the economics literature document large differences in labour market outcomes across college majors. Altonji et al. (2012) review the literature on the returns to college curriculum and show that the evidence on the heterogeneity of returns across majors has remained remarkably consistent over time, with some majors such as engineering commanding a high premium and others including humanities, social sciences and education further behind.

The sorting of women in less remunerative fields has been investigated as one of the factors accounting for the gender gap in earnings. Flabbi (2012) examines the impact of educational choices of females and males on their respective labour market outcomes for 14 OECD countries, and demonstrates that, when not controlling for job characteristics, gender differences in the field of study explain approximately 16% of the gender gap in earnings. Moreover, he shows that the returns to the field of study are different between females and males and that this difference is the most important component of the overall unexplained part of the earnings gap.[2] Card and Payne (2017) focus on differences in graduation rates in STEM (science, technology, engineering and mathematics) fields and illustrate that these explain between 1/5 and 1/10 of the wage gender gap among Canadian full time workers.

The evidence on the role of the field of study in explaining the gender earnings gap of Italian workers is mixed: Anelli and Peri (2015a) analyse a sample of individuals from high quality college preparatory high schools in a large city in northern Italy and find that up to one third of the gender gap in earnings is attributable to the choice of major. Their evidence is contrasted by results from Piazzalunga (2018), illustrating that in a sample consisting of one cohort of Italian college graduates at the beginning of their careers, the inclusion of academic variables (including field of study) in the wage equation does not reduce the magnitude of the gender gap

[2]He finds that, for males, choosing any field which is not humanities increases the wage of a significant amount, while, for females, only graduating from social sciences significantly increases the wage relative to humanities.

coefficient remarkably.

I contribute to this literature by extending the analysis to a bigger sample, which covers 65% of the entire population of college graduates from the cohorts 2010 to 2012. More importantly, thanks to the richness of the dataset used that I complement with administrative data on the supply of higher education from the Italian Ministry of Education (MIUR), I am able to characterise precisely the content of the specific degree course from which students in the sample graduated.

Furthermore, I add an important element to the analysis of the gender earnings differences, by being able to investigate the role of aspects related to individuals' personal traits. Recent studies have started analysing the difference in psychological traits and preferences between females and males. The findings from these studies indicate that women are more risk averse and less willing to compete, more socially minded and more altruistic (see Booth and Nolen (2009), Gneezy et al. (2003), Niederle et al. (2013), Andreoni and Vesterlund (2001), Eckel and Grossman (1998)). Most of the evidence on the gender differences in personal traits comes from experimental settings, while the evidence on their impact on labour market outcomes is less rich. One example is Fortin (2008), who investigates the impact of non-cognitive traits – including the importance of money/work and the importance of people/family – on wages and on the gender wage gap among young workers and finds that these traits have a significant, although modest, role in accounting for the gender wage gap. I contribute to this scarce literature by investigating the role of psychological traits in accounting for gender differences in labour market outcomes. I am able to extract information on workers' personal traits through answers to questions related to preferences for different aspects of a job.

I estimate a wage equation including a rich set of variables: demographic variables; variables measuring human capital accumulated through education, from high school to college; measures of the socio-economic background; and measures of personality traits. The raw gender gap in the average monthly wages in my final sample is 25.7%, going down to 14.6% when restricting to full-time workers. The results of my analysis show that the variables related to the content of the degree course play the biggest role in accounting for the gender difference in earnings conditional on full-time status. Approximately one fourth of the gender gap in monthly wages three years after graduation is explained by gender differences in the field of study at college. Most interestingly, I find that this result can be attributed to a specific feature of the university courses that females are less likely to choose, which is the maths content. Females sort into degree courses with a lower maths content, which are also the highly remunerative ones. Furthermore, I find that differences in psychological traits have a modest but significant role in explaining the gender gap in earnings. On the other hand, the results indicate a negligible role of family characteristics and high school choices, over and above effect they can already have on the major choice.

The remainder of the chapter is organised as follows. Section 1.2 presents the uniquely rich dataset on college graduates used for the analysis, and gives details on the characteristics of the final sample. The empirical methodology used for analysing the gender earnings gap is described in Section 1.3. Section 1.4 presents and discusses the results from the wage equation estimation and the Oaxaca decomposition of the female-male earnings differential. Section 1.5 concludes the chapter.

1.2 Data and Final Sample

1.2.1 The AlmaLaurea dataset

In order to analyse the determinants of the gender earnings gap among recent college graduates, I exploit a uniquely rich and largely unexplored dataset provided by AlmaLaurea, an inter-university consortium collecting data on students who graduate from the universities that are part of the consortium.

AlmaLaurea's original institutional objectives are twofold: first, to provide member academic institutions with reliable information on their students by managing a database that collects information on graduates; second, it aims at facilitating the graduates' labour market transition by managing a service that gives firms electronic access to graduates' curriculum vitae.

Data on graduates are drawn from two different sources: first, academic institutions provide official data on students' demographic information and on their university careers. The administrative variables originated from this source are: students' date of birth, municipality of birth and of residence at time of university enrolment, high school attended and final grade, year and course of enrolment in university, university GPA, date of discussion of the dissertation and graduation grade. Second, upon graduation students complete a survey providing several pieces of information, among which: family characteristics, satisfaction from the university experience, level of other skills including language and IT skills, study experiences abroad, other training experiences, intention to continue studies, and aspirations about the future career. All these variables form the dataset referred to as *Graduates' Profile*. The historical series of this survey contains data on graduates' cohorts from 2004 to 2015.

With the goal of monitoring graduates' access to the labour market, AlmaLaurea follows graduates one, three and five years after graduation. The survey is entitled *Graduates' Employment Conditions* and provides information on: graduates' employment status, time span between graduation and first job, effectiveness of the degree for finding a job, characteristics of the current job including salary, type and location of job, and satisfaction with the job. Graduates with an undergraduate degree are interviewed only one year after graduation, and, in case they pursue a master's degree, again at graduation and 1, 3 and 5 years after graduation.

Participation in the survey from universities is voluntary: it implies the payment of a one-off membership fee and a yearly payment proportional to the total number of graduates, in exchange for the services provided by the consortium. Throughout the years more universities progressively took part in the survey. I will focus on students who graduated from 2010 to 2012 from the 56 universities surveyed every year in the period considered. The Italian higher education system in this period was composed of 89 institutions[3], including 11 long-distance-learning institutions, 3 universities for foreigners and 75 traditional universities, both public and private. Figure 1.2 illustrates the geographical distribution of the Italian universities (excluding the long-distance-learning institutions) highlighting those that are in the AlmaLaurea sample. Some important institutions are not part of the sample in the period considered: namely, the two most important state universities, the technical university and the two major private universities in a major city in the north-east of the country (Milan);

[3] Excluding one institution accredited in 2011.

the biggest university in a major city in southern Italy (Naples); and a very important university in Sicily. In table 1.1 I report the distribution of the universities in the population and in the AlmaLaurea sample across various dimensions. It can be noticed that there are no significant differences in terms of size of the universities or field of study of the courses offered by the institutions. The AlmaLaurea sample contains no long-distance-learning institutions, while public universities are more represented.

Overall across all cohorts the AlmaLaurea sample covers approximately 65% of the population of the Italian college graduates; table 1.2 reports the distribution of students across fields of study by gender in the population and in the sample, and demonstrates that the two distributions are very close.

Once a university takes part in the consortium, it provides administrative information on the universe of its graduates. Response rate to the questionnaire at graduation is very high: between 91 and 93% of students complete the survey each year. Three years after graduation the response rate is still remarkably high, ranging between 74 and 80%. In table 1.3 I report the response rate at graduation and three years after, by graduation cohort and type of degree.

For the purpose of my analysis I use administrative and survey data from AlmaLaurea *Graduates' Profile* on master and single-cycle college graduates from cohorts 2010-2012, combined with data on employment status and earnings three years after graduation from the Almalaurea *Graduates' Employment Conditions*. From the administrative variables I take demographic information – i.e., gender, year and municipality of birth – and information on the educational path, from high school – i.e., high school track[4], institution attended and final grade – to college – i.e., university attended, degree course, performance measured by GPA, final graduation grade, and experiences of study abroad. From students' answers to the questionnaire I extract other variables, namely: other skills including number and level of knowledge of foreign languages and number of IT tools in which they are skilful; family

[4] In Italy, the secondary education system is organized in several different study paths. Students can choose among: a 'scientific' high school offering students a maths- and science-intensive curriculum; humanities-intensive high schools including 'classics', 'education', 'languages' and 'artistic' tracks; 'technical' high schools offering specialisation in technological subjects, either with a focus on business, tourism or agriculture (non-STEM) or with a focus on industrial construction and preparation for surveyors (STEM).

characteristics, i.e., the level of education of father and mother and their last occupation; students' preferences for the future work career, through answers to questions about how much they value aspects of the job such as salary and career prospects, adherence to cultural interests, stability of the job position and availability of free time; their employment status three years after graduation and, if employed, their full/part time status, monthly wage and the location of the job (Italian province).

1.2.2 Final sample: selection and summary statistics

The overall number of masters' and single-cycle college graduates from 2010-2012 cohorts is approximately 220,000, of which 71% are interviewed both at graduation and 3 years later. I focus on individuals born in Italy and residing in Italy upon graduation and who graduated between 23 and 31 years old, excluding 8% of the observations. 66% of these individuals are employed three years after graduation. The final sample is made by 71,220 employed workers for whom there is information on all the variables of interest, of which 76% (53,851) are full time workers.

Table A1 in the appendix summarises the effects of sample selection on the characteristics of the final sample. The first important selection is based on the response to the surveys, both upon graduation and three years after. Administrative variables – i.e., gender, age, university and high school career variables – are available for the entire population of college graduates from the 56 universities surveyed by AlmaLaurea (column (1)). Females are the majority of the population of college graduates (60%). Respondents to the surveys (column (2)) do not appear to be selected according to any of these variables. The sample of interviewed students selected based on place of birth and residence and age (column (3)) is not significantly different from the initial population in any of the administrative variables, and from the sample of respondents in any of the survey variables – i.e., measures of skills, family characteristics and preferences.

The second important selection is in excluding individuals who are not employed, both the ones looking for a job and the ones not participating to the labour market (respectively 19% and 16%).[5] The characteristics of the sample of employed

[5]The AlmaLaurea definition of employed workers excludes individuals who are undergoing some academic or professional training (post-graduate courses including PhDs, internships and trainee-

individuals are presented in column (4). Females are only slightly less represented, indicating that labour market attachment among recent college graduates is not dramatically different between females and males. The employment gender gap in my sample is of 7.7 percentage points, less than half of the gender gap in employment measured by the Italian National Institute of Statistics (ISTAT) for all levels of education, and lower than the same measure for all individuals (not only recently graduated) with at least a college degree (approximately 10 percentage points in the years 2013-2015). Employed individuals are more likely to have graduated from an engineering course, while health graduates are less represented relative to the initial sample; this last result is most likely driven by medicine graduates undergoing residencies, who, according to the AlmaLaurea definition of employment status, are considered as non-employed. Among employed individuals the distribution across high school tracks is slightly changed: individuals who completed a technical high school – in particular offering preparation in STEM fields – are more represented, while the opposite is observed for individuals who attended the high school track focused on classics. Employed individuals are also slightly better selected in terms of IT skills, but negatively in terms of GPA. Finally, it seems that preferences are related to the employment status, with employed people valuing more career prospects and less cultural interests and free time. In the sample of full-time workers (column (5)) the selection based on gender and field of study is even stronger: females are only 52% of this group of workers, and engineering graduates are more represented while there are less graduates from education, humanities and social sciences.

Overall, the selection bias based on observables does not have a clear direction: employed individuals are better selected in some characteristics, but other variables suggest they are endowed with worse skills. It is worth noticing that characteristics related to the preparation in science and technology, from the the high school track to the field of study in college and the IT skills, appear to be positively correlated with being employed.

In order to investigate whether there are strong differences in the selection into

ships, residencies for medicine graduates), even if paid, contrary to the definition of the Italian National Institute of Statistics (ISTAT) that includes this group in the employed population. In the AlmaLaurea sample, this group of workers represents approximately 70% of all the individuals who are not working and not looking for a job.

employment across genders, I look at the characteristics of employed and not employed individuals in the samples of females and males, that are presented in table A2 in the appendix. Men are much more likely to work, and to work full time (86% vs 68% for women). For both genders, employed individuals are more likely to come from engineering and less from social science, science and maths; they have slightly higher GPA in college and better IT skills and are positively selected in terms of socio-economic background. Some differences between the two sexes emerge: (i) men who studied humanities are less likely to be employed but this is not observed for females; (ii) men who completed a technical STEM high school are more represented in the sample of employed, but this is not true for females; (iii) contrary to what expected, unemployed females have higher preferences for salary and career aspects.

The final samples are composed by employed workers or full-time only workers for whom there is information on all the variables of interest (respectively columns (6) and (7) of table A1). Table 1.4 presents summary statistics for both samples, separately for females and males. At university females are more represented in education, social sciences and humanities, while less in science, maths and especially in engineering; at high school they are more likely to have completed humanities rather than science-intensive and technical STEM tracks. On average, women have better college and high school performances, they are more likely to undergo post-graduate training and to have done experiences abroad during college (in particular in the full-time workers sample), they speak more languages, but have lower knowledge of IT tools. On the other hand, men have slightly better family characteristics. Finally, females give lower importance to career prospects and higher importance to job stability and adherence of the job to cultural interests.

Overall, it emerges that females, on average, are endowed with better skills, but they accumulate less human capital related to science and technology, and they are endowed with 'soft' skills related to lower competitiveness and higher social mindedness.

1.3 Methodology

I estimate a wage equation in which wage is function of demographic variables, human capital measures and measures of preferences, as well as family characteristics. I estimate the following specification:

$$Y = F\beta_1 + X\beta_2 + H\beta_3 + P\beta_3 + S\beta_4 + u \qquad [1.1]$$

where Y is average monthly wage of each student, F is a binary variable taking value 1 for female students, X is a vector of demographic control variables, H is a vector of variables measuring human capital, P is a vector of variables measuring preferences and S is a vector of socio-economic background variables. All the parameters are estimated through OLS; the parameter of interest, β_1, identifies the conditional wage differentials between men and women when controlling for other independent variables.

In order to control for gender differences in returns to the different characteristics, I perform an Oaxaca decomposition of the earnings gender gap, which decomposes the estimated female-male difference in earnings in a part that is 'explained' by group differences in characteristics and a part given by differences in the returns to the same characteristics. The difference in the expected value of the outcome variable Y among females and males is implemented in the following way:

$$E(Y_F) - E(Y_M) = \{E(Z_F) - E(Z_M)\}' \gamma_M + E(Z_F)'(\gamma_F - \gamma_M) \qquad [1.2]$$

where Z denotes a vector containing all the predictors and a constant and γ contains the slope parameters and the intercept. The difference in characteristics is weighted by males coefficients, while the difference in coefficients is weighted by females characteristics.

1.3.1 Content of degree courses

Exploiting the richness of the AlmaLaurea dataset, I am able to control for detailed variables related to what each student in the sample studied in college. College majors can be classified in 26 broad fields of study according to the OECD classification of *Fields of Education and Training - Foet1999*. Within each field, smaller groups of degree courses can be distinguished, indicated as 'classes' of degree. Most importantly, AlmaLaurea records from the different institutions the name of the precise degree course from which the student graduated. I complement this information with a unique dataset made available by the Italian Ministry of Education (MIUR) on the supply of higher education in Italy, to characterise precisely the content of the course each student graduated from. In particular, the dataset contains a list of all courses offered by each single university each year (since 2001), and for each of them it provides detailed information on the content in terms of subjects studied. The information on the content comes in form of the number of credits in the European Credit Transfer System (ECTS) that the students have to be awarded in each of 370 different 'disciplinary sectors'.

I use this information to characterise each degree course with an index indicating the intensity of the maths content. This choice is motivated by the idea that difference in returns across college majors can be attributed to difference in maths ability. Some evidence in this direction comes from Paglin and Rufolo (2016), who show that 82% of the variance across college majors in entry-level wages is explained by the average GRE-maths scores by major. They show that fields with a high proportion of women are lower paying because the human capital in these fields can be produced with "*less of an important scarce attribute (quantitative ability)*", and vice-versa.

Hence, I classify each of the 370 disciplinary sectors as maths-intensive or non maths-intensive and I construct a maths intensity index, which is the proportion of maths-intensive credits out of all credits for each course. This index is obtained for more than 4,000 unique undergraduate, master and single-cycle university courses offered by single higher education institutions each year. Figure 1.3 illustrates the average and the standard deviation of the index across all courses offered in 2010 within each of the 26 broad FOET1999 fields of study. There is a lot of heterogeneity across fields of study: courses in humanities and education have on average maths content close to zero, while for courses in maths & stats, physics or engineering almost the totality of subjects studied is maths-intensive. The standard deviations indicate that even within each field there is a high level of heterogeneity across different specific degree courses.

I will merge these administrative data from the MIUR with the Almalaurea dataset to characterise the degree course attended by each student in its maths content.

1.4 Results

Table 1.5 reports the female coefficients estimated in different specifications where the included types of controls vary. The unadjusted gender gap in average monthly wages is 25.7% (column (1)), which does not change considerably when controlling for demographic variables including municipality of birth and graduation cohort and for province of job (column (2)). This value is lower than other estimates of the gender earnings gap in Italy: for the same period (2013 to 2015), the Global Gender Gap reports indicate that overall in the full population of workers the gender gap in earnings (women-men difference as ratio of men earnings) was, on average, 43%; Anelli and Peri (2015a) find that the annual earnings of female college graduates in their 30's and 40's observed between 5 and 15 years after graduation were 37% lower than the ones of males. The difference with this evidence is driven by the fact that my results are obtained for a very homogeneous sample of workers in terms of educational attainment, age, and potential labour market experience. Despite this, I still find a sizeable gender gap: highly educated female workers at the beginning of their careers already earn considerably less than their male peers. This result is in contrast with other evidence showing that the wage gap is small upon entrance in the labour market and builds up later in life, especially because of lower hours worked mainly attributable to career interruptions for childrearing (see for example Bertrand et al. (2010) for US).[6]

Controlling for full-time status, as expected, makes the coefficient drop significantly to 13.7% (column (3)).

In columns (4) and (5) high school controls – grade and curriculum – are added:

[6]Bertrand et al. (2010) find that MBA graduates in the US show no wage gaps upon graduation, but large gaps build over first 10 years of labour market experience, mostly due to the presence of children.

the coefficient of interest drops by approximately 8%, and slightly more when high school fixed effects are included. Because females perform better in high school with respect to their males peers, this effect must be driven by differences in the high school curriculum chosen. Since here we are not controlling for academic variables, it has to be considered that part of this effect of the high school experience on earnings could be mediated through an effect on college choice. Academic variables measuring performance at university and level of other skills and human capital accumulated do not account for much of the earnings gap (column (6)).

Specifications in columns (7), (8) and (9) control for the subjects studied at college, respectively from the broad 26 fields of study to the 100 classes of degree and finally the approximately 3,300 different specific university courses offered. Variables related to the college major produce a significant drop in the gender gap coefficient, from 12% up to 25% when the courses fixed effects are included. Differential sorting of females and males in college courses is an important determinant of the gender gap in earnings. The most interesting result is that approximately the same drop is produced when controlling for the maths intensity of the 3,300 university-degree courses (column (10)), suggesting that the characteristic of degree courses relevant for explaining gender differences in earnings is its maths content.

In the two final columns I add respectively controls for family characteristics and variables measuring preferences: with the former, which proxy for socio-economic status, the coefficient of the gap barely changes, indicating a small role of the family influence on earnings over and above the impact it can already have on high school and college choices. Preferences have a modest impact on the gender gap coefficient; this result may suggest that the effect of these 'soft skills' on labour market outcomes goes through their impact on educational choices, while the effect on earnings on top of the educational choices is small.

Results from the same estimations implemented for the sample of full-time workers lead to the same conclusions (table 1.6).

Taking into consideration that controlling for high school fixed effects, as well as for degree courses fixed effects, does not make the coefficient of interest significantly change, my preferred specifications include high school curriculum in 8 categories, and subjects studied at college respectively as 100 classes of degrees or as maths content index. Consequently, I will perform an Oaxaca decomposition of the female coefficient estimated in these two specifications, both for the full sample and the full-time workers sample.

Results from the Oaxaca decomposition are reported in table 1.7. For both specifications in both samples, the gender earnings gap is accounted for in approximately the same proportion by differences in characteristics and differences in coefficients. Columns (1) and (3) indicate that the variables related to the subjects studied at college and the preferences constitute the bigger portion of the part of the gap explained by differences in characteristics. Females sort in degree courses with lower returns on the labour market, and have preferences for aspects of the job that are negatively associated to higher earnings. In particular, the results from the Oaxaca decomposition performed on the specification that includes the maths content of degree courses (columns (2) and (4)) indicate that the female-male difference in maths intensity of the college course attended accounts for approximately 15% of the earnings gap, up to 27% in the full-time workers sample. When looking at the coefficients terms, the results indicate that the only relevant factor is the difference in returns to the maths content of the courses: even conditional on the maths intensity of the degree course chosen, females have much lower returns on the labour market.

The results of my analysis indicate that, even in a sample of workers relatively homogeneous in terms of potential work experience, age and human capital, I am able to explain a considerable part of the remaining gender gap in earnings. I am able to control for detailed variables not available in common surveys measuring factors that other studies failed to take into account, in particular detailed human capital variables generated very early in life – from high school choices made at age 14 to college major choices made at 18 years old – and variables related to personality traits.

1.5 Concluding Remarks

This study analyses the gender gap in earnings among Italian recent college graduates who are at the beginning of their career, controlling for a rich set of variables measuring students' educational experience, their socio-economic background and aspects of their personalities.

Findings indicate that in Italy there is a significant gender gap in monthly wages among college graduates already three years after graduation, despite the fact that among this highly educated group women do not show a significantly lower attachment to labour market relative to men. Moreover, these workers are at the beginning of their careers, and explanations as career interruptions due to childrearing do not apply yet.

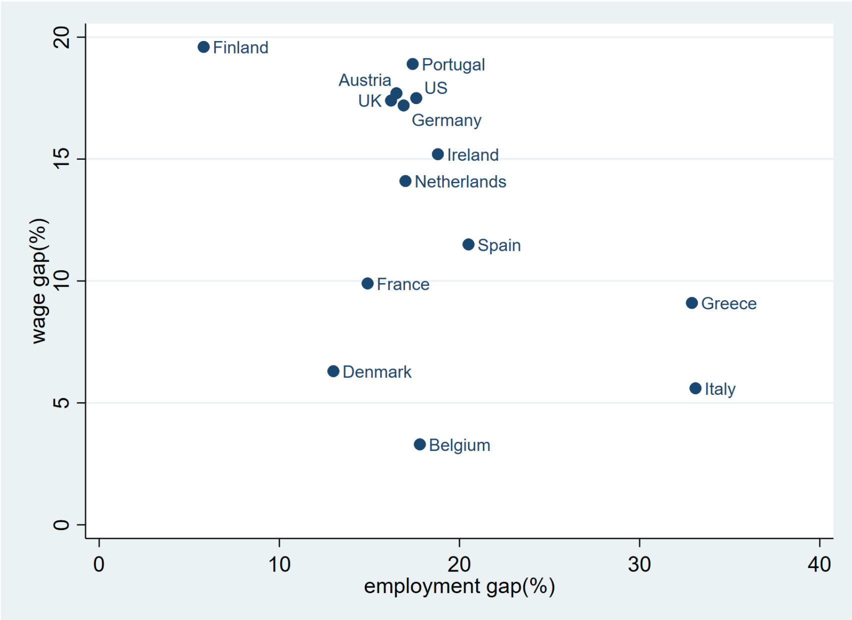
Thanks to the richness of the dataset used, I am able to control for a set of variables that are not available in common surveys and were before omitted from wage equation estimations, in particular related to the content of the college degree course and aspects related to personality.

The study shows that women are better endowed in terms of human capital, i.e., they perform better both in high school and college and tend to have higher level of other skills, but this does not translate in an advantage in the labour market. On the other hand, females have characteristics negatively associated to wages. In particular, they graduate less from maths-intensive high remunerative fields and are characterised by personal traits negatively associated with future wages. Even conditional on graduating from a degree course with high maths content, women have much lower returns to this choice.

By showing that college major is the most significant variable in explaining gender earnings gap, my results suggest the importance of investigating more in depth the forces driving differences in educational choices in college between female and male students.

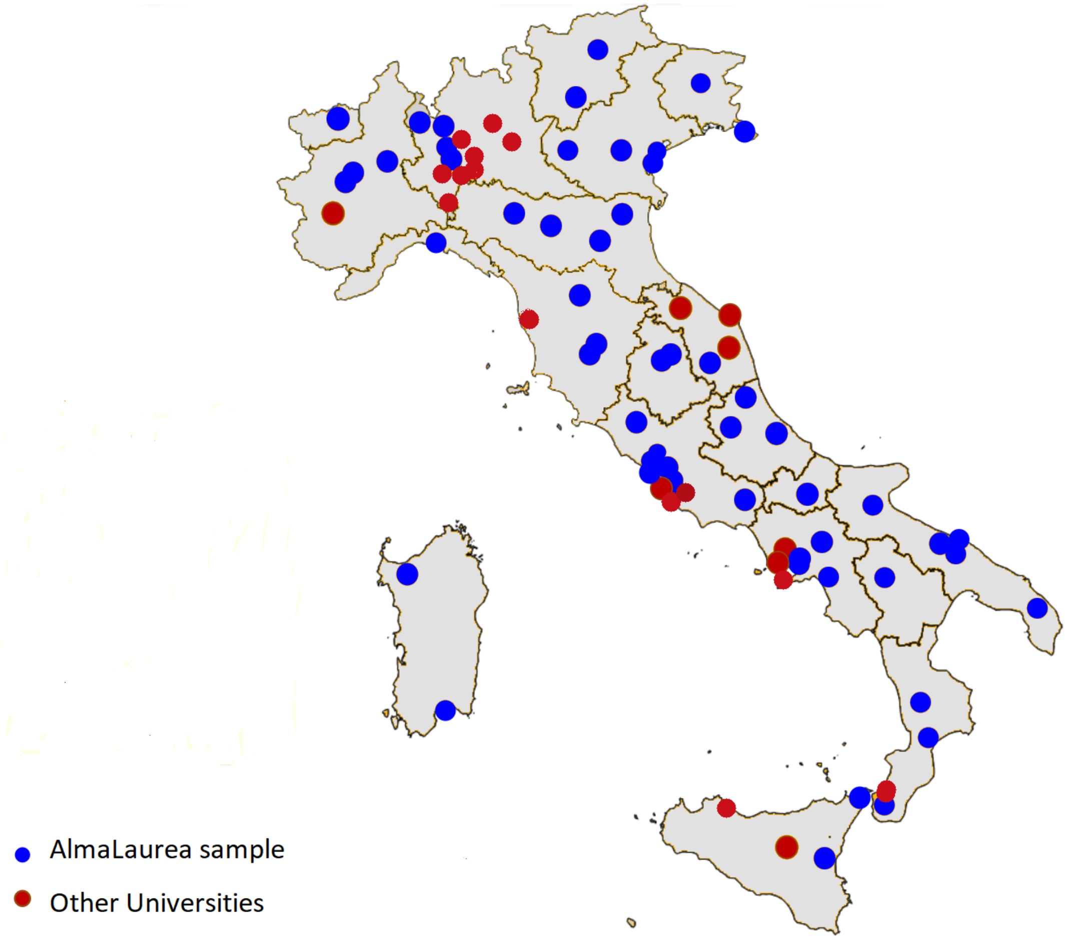
Figures and Tables

Figure 1.1: Gender pay and employment gap across EU-14 countries



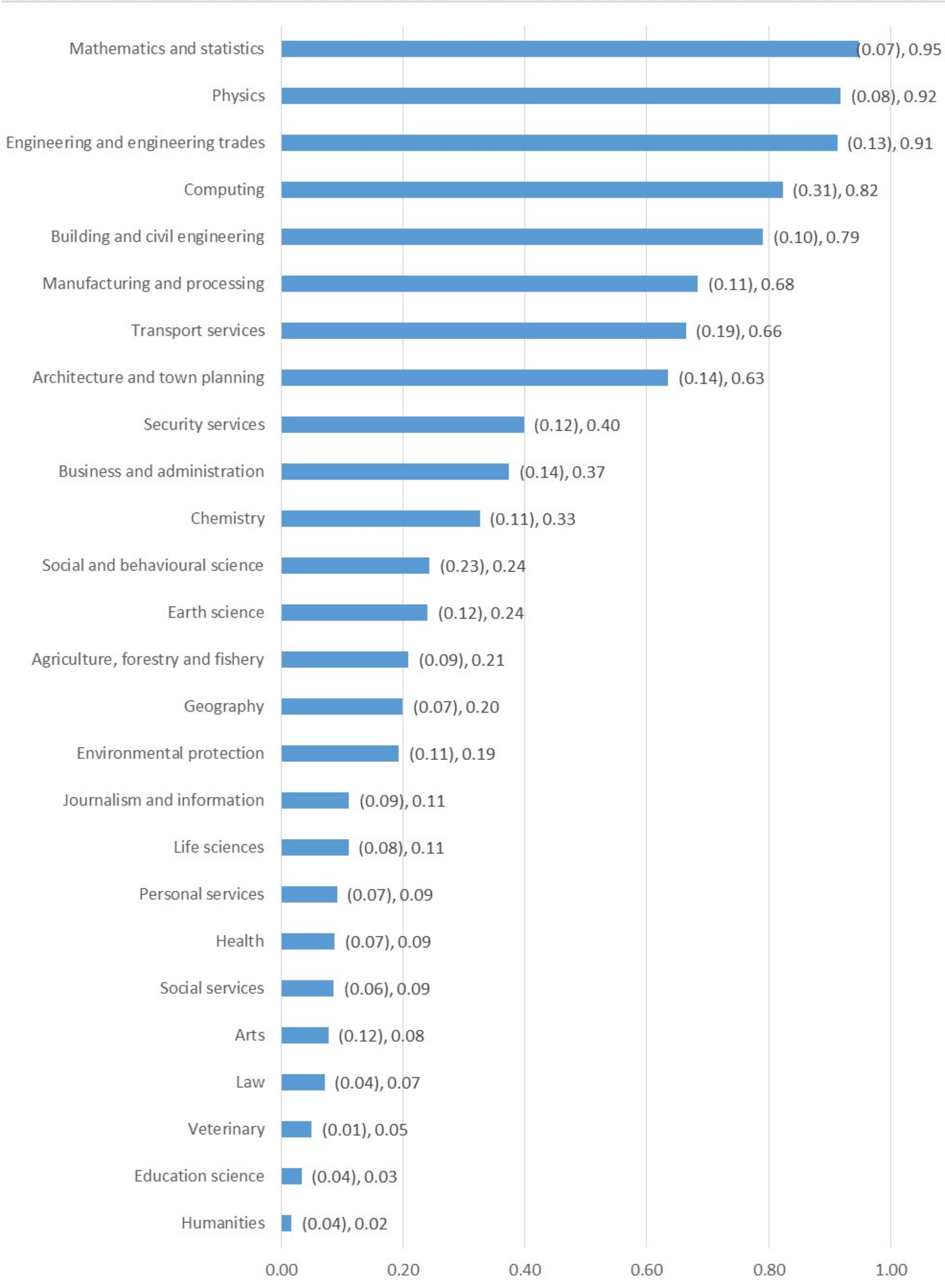
Notes: The figure plots the employment gender gap and gender wage gap for 14 OECD countries. The gender wage gap is unadjusted and is defined as the difference between median annual earnings of men and women relative to median annual earnings of men. Data refer to full-time employees and self-employed. Source: OECD (2015a).

Figure 1.2: Map of the Italian higher education system



Notes: The figure plots the 78 (non long-distance-learning) Italian higher education institutions existing in 2015, by geographical location and distinguishing those not surveyed by AlmaLaurea.

Figure 1.3: Fields of study and their maths content



Notes: Average maths content index across all degree courses within each of the 26 FOET1999 fields of study. Standard deviations are reported in parentheses.

Table 1.1: AlmaLaurea Sample: Universities

Distribution of Universities (% over total)

	AlmaLaurea sample	All Universities
Size (n. students)		
<10000	41.07	43.82
10000-20000	21.43	22.47
20000-40000	25	21.35
>40000	12.5	12.36
Type		
Long-Distance-Learning	0	12.4
Private	8.93	11.2
Public	91.07	76.41
Courses offered by field		
Education	9.6	9.86
Humanities and Arts	14.9	14.08
Social sciences, business and law	16.56	18.08
Science, Maths and Computing	13.58	12.44
Engineering, Manufacturing	12.91	12.68
Agriculture	7.28	7.28
Health and Welfare	13.25	13.15
Services	11.92	12.44

Notes: Data on the population of Italian universities and number of graduates are taken from the Office of Statistics of the Italian Ministry of Education.

Table 1.2: AlmaLaurea Sample: Students

Distribution of students by gender and field of study (% over total)

Field of study	AlmaLaurea sample			All Universities		
	Males	Females	All	Males	Females	All
Education	0.8	6.0	3.9	0.8	5.8	3.7
Humanities and Arts	9.5	20.3	16.0	8.3	19.2	14.7
Social sciences, business and law	32.2	34.7	33.7	34.9	35.8	35.4
Science, Maths and Computing	10.2	7.7	8.7	9.4	7.9	8.5
Engineering and Manufacturing	29.2	9.9	17.6	29.3	10.4	18.3
Agriculture	2.7	1.6	2.0	2.6	1.6	2.0
Health and welfare	11.8	17.4	15.1	10.6	16.7	14.1
Services	3.7	2.5	3.0	4.0	2.6	3.2

Notes: Data on the population of Italian graduates are taken from the Office of Statistics of the Italian Ministry of Education.

Table 1.3: Response Rate by Graduation Cohort and Type of Degree

Panel A: Response rate at graduation (%)

Year of graduation	Type of degree			
	Undergraduate	Single cycle	Master	Total
2010	92	90	91	91
2011	94	91	92	93
2012	93	91	90	92
Total	93	91	91	92

Panel B: Response rate three years after graduation (%)

Year of graduation	Type of degree			
	Undergraduate	Single cycle	Master	Total
2010	-	78	80	80
2011	-	76	76	76
2012	-	74	75	74
Total	-	76	77	77

Notes: The sample consists of all college graduates from the 56 universities surveyed by AlmaLaurea every year from 2010.

Table 1.4: Summary Statistics

	(1)	(2)	(3)	(4)		(1)	(2)	(3)	(4)
VARIABLES	Males	Females	FT Males	FT Females	VARIABLES	Males	Females	FT Males	FT Females
Age	29.43	29.14	29.39	29.04	**High School:**				
	(1.786)	(1.713)	(1.764)	(1.673)	Final grade 60-84	0.508	0.379	0.499	0.363
Full time	0.859	0.684				(0.500)	(0.485)	(0.500)	(0.481)
	(0.348)	(0.465)			Final grade 84-95	0.235	0.263	0.240	0.262
Monthly wage	1,312	1,042	1,402	1,213		(0.424)	(0.440)	(0.427)	(0.440)
	(555.2)	(490.3)	(514.0)	(438.9)	Final grade 95-100	0.257	0.358	0.261	0.376
Field of Study						(0.437)	(0.480)	(0.439)	(0.484)
Education	0.00282	0.0300	0.00218	0.0202	*Curriculum:*				
	(0.0531)	(0.170)	(0.0466)	(0.141)	Classics	0.107	0.197	0.0972	0.186
Humanities and Arts	0.0665	0.186	0.0487	0.162		(0.309)	(0.397)	(0.296)	(0.389)
	(0.249)	(0.389)	(0.215)	(0.368)	Education	0.00677	0.0883	0.00554	0.0671
Social sciences, business and law	0.335	0.417	0.336	0.421		(0.0820)	(0.284)	(0.0743)	(0.250)
	(0.472)	(0.493)	(0.472)	(0.494)	Languages	0.0177	0.113	0.0157	0.113
Science, Maths and Computing	0.0856	0.0708	0.0855	0.0734		(0.132)	(0.317)	(0.124)	(0.316)
	(0.280)	(0.256)	(0.280)	(0.261)	Art	0.00810	0.0180	0.00737	0.0163
Engineering and Manufacturing	0.385	0.141	0.421	0.173		(0.0896)	(0.133)	(0.0855)	(0.127)
	(0.487)	(0.348)	(0.494)	(0.378)	Technical non-STEM	0.116	0.145	0.117	0.159
Agriculture	0.0294	0.0232	0.0290	0.0252		(0.320)	(0.352)	(0.322)	(0.366)
	(0.169)	(0.151)	(0.168)	(0.157)	Technical STEM	0.205	0.0226	0.215	0.0252
Health and Welfare	0.0778	0.119	0.0649	0.113		(0.404)	(0.149)	(0.411)	(0.157)
	(0.268)	(0.324)	(0.246)	(0.317)	Scientific	0.526	0.405	0.528	0.423
Services	0.0178	0.0135	0.0134	0.0114		(0.499)	(0.491)	(0.499)	(0.494)
	(0.132)	(0.115)	(0.115)	(0.106)	Professional	0.0138	0.0108	0.0136	0.0101
Academic variables:						(0.117)	(0.104)	(0.116)	(0.0998)
Graduation grade	106.5	108.0	106.4	107.9	**Family characteristics:**				
	(6.656)	(5.940)	(6.659)	(5.997)	*Father education:*				
Late degree (index)	0.269	0.233	0.268	0.223	Less than high school	0.300	0.353	0.301	0.344
	(0.382)	(0.362)	(0.381)	(0.356)		(0.458)	(0.478)	(0.459)	(0.475)
GPA 18-26	0.422	0.309	0.428	0.321	High school	0.443	0.437	0.446	0.441
	(0.494)	(0.462)	(0.495)	(0.467)		(0.497)	(0.496)	(0.497)	(0.496)
GPA 27-29	0.443	0.486	0.448	0.488	College degree	0.171	0.140	0.166	0.141
	(0.497)	(0.500)	(0.497)	(0.500)		(0.377)	(0.347)	(0.372)	(0.348)
GPA 29-30	0.135	0.204	0.124	0.191	College degree science and engineering	0.0861	0.0697	0.0873	0.0745
	(0.341)	(0.403)	(0.330)	(0.393)		(0.281)	(0.255)	(0.282)	(0.263)
Other training ongoing	0.101	0.138	0.0670	0.0808	*Mother education:*				
	(0.302)	(0.345)	(0.250)	(0.272)	Less than high school	0.302	0.340	0.304	0.332
Other training completed	0.522	0.562	0.538	0.603		(0.459)	(0.474)	(0.460)	(0.471)
	(0.500)	(0.496)	(0.499)	(0.489)	High school	0.479	0.473	0.481	0.477
Number of foreign languages:						(0.500)	(0.499)	(0.500)	(0.499)
0	0.273	0.257	0.270	0.245	College degree	0.177	0.152	0.173	0.154
	(0.446)	(0.437)	(0.444)	(0.430)		(0.382)	(0.359)	(0.378)	(0.361)
1	0.538	0.422	0.548	0.424	College degree science and engineering	0.0415	0.0355	0.0417	0.0374
	(0.499)	(0.494)	(0.498)	(0.494)		(0.200)	(0.185)	(0.200)	(0.190)
2	0.167	0.259	0.162	0.267	*Social class:*				
	(0.373)	(0.438)	(0.368)	(0.443)	Managerial and professional workers	0.271	0.235	0.272	0.244
3	0.0204	0.0589	0.0188	0.0599		(0.445)	(0.424)	(0.445)	(0.430)
	(0.142)	(0.235)	(0.136)	(0.237)	Intermediate occupations	0.319	0.317	0.319	0.317
4	0.00143	0.00328	0.00119	0.00381		(0.466)	(0.465)	(0.466)	(0.465)
	(0.0378)	(0.0571)	(0.0344)	(0.0616)	Non professional self employed	0.192	0.222	0.193	0.226
IT skills (number of IT tools):						(0.394)	(0.416)	(0.394)	(0.418)
0	0.0160	0.0292	0.0142	0.0253	Routine work	0.217	0.226	0.216	0.212
	(0.126)	(0.168)	(0.118)	(0.157)		(0.413)	(0.418)	(0.412)	(0.409)
1-2	0.0807	0.149	0.0722	0.137	**Preferences:**				
	(0.272)	(0.356)	(0.259)	(0.344)	Importance salary	0.546	0.530	0.555	0.537
3-4	0.276	0.383	0.272	0.387		(0.498)	(0.499)	(0.497)	(0.499)
	(0.447)	(0.486)	(0.445)	(0.487)	Importance career prospects	0.681	0.587	0.701	0.612
5 or more	0.627	0.439	0.641	0.451		(0.466)	(0.492)	(0.458)	(0.487)
	(0.484)	(0.496)	(0.480)	(0.498)	Importance job stability	0.581	0.692	0.577	0.685
Period abroad: Erasmus	0.142	0.143	0.144	0.156		(0.493)	(0.462)	(0.494)	(0.464)
	(0.349)	(0.350)	(0.351)	(0.362)	Importance culture	0.368	0.458	0.349	0.436
						(0.482)	(0.498)	(0.477)	(0.496)
					Importance free time	0.213	0.219	0.203	0.207
						(0.409)	(0.414)	(0.402)	(0.405)
					Observations	29,399	41,821	25,251	28,600

Notes: The sample consists of masters' and single-cycle college graduates from 2010-2012 cohorts, born in Italy and residing in Italy upon graduation, who graduated between 23 and 31 years old, and who are employed three years after graduation. Columns (3) and (4) refer to full time workers only.

Table 1.5: Female Coefficients from OLS Regressions

	Baseline			With high school controls		With academic and major controls					With family char.	With preferences
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Female	-0.257***	-0.241***	-0.137***	-0.126***	-0.123***	-0.113***	-0.0987***	-0.0848***	-0.0844***	-0.0981***	-0.0848***	-0.0814***
	(0.00442)	(0.00427)	(0.00365)	(0.00406)	(0.00469)	(0.00463)	(0.00458)	(0.00465)	(0.00486)	(0.00476)	(0.00483)	(0.00487)
Controls:												
Demographic controls		x	x	x	x	x	x	x	x	x	x	x
Full time dummy			x	x	x	x	x	x	x	x	x	x
High school fixed effects					x	x	x	x	x	x	x	x
26 fields of study							x					
100 groups of degrees								x				
Course-university fixed effects									x		x	
Maths content of course-university										x		
Family characteristics											x	x
Observations							71,219					
R-squared	0.050	0.230	0.416	0.421	0.501	0.511	0.554	0.563	0.593	0.514	0.593	0.594

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Notes: The sample consists of masters' and single-cycle college graduates from 2010-2012 cohorts interviewed both at graduation and 3 years later, born in Italy and residing in Italy upon graduation and who graduated between 23 and 31 years old, who are employed 3 years after graduation and for whom there is info on all the variables of interest. Demographic controls include municipality of birth, year and age at graduation, and province of job. High school controls include the final grade and the track in 8 categories (column (4)) or school-track dummies (column (5)). Academic controls include college final grade, GPA and time for completion, experiences abroad, language and IT skills, post-graduate training and 26 fields of study (column (7)) or 100 classes of study (column (8)) or 3,300 university-degree courses dummies (column (9)) or maths content index for 3,300 university-degree courses (column (10)). Family characteristics are the level of education of father and mother and the social class of the family (column (11)). Preferences are the variables taking value 1 if the student indicates as very important for the future job the following aspects: salary, career prospects, culture, stability, and free time for job (column (12)).

Table 1.6: Female Coefficients from OLS Regressions (Full-time Workers Sample)

	Baseline		With high school controls		With academic and major controls					With family char.	With preferences
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Female	-0.146***	-0.149***	-0.136***	-0.133***	-0.121***	-0.101***	-0.0869***	-0.0861***	-0.100***	-0.0863***	-0.0814***
	(0.00423)	(0.00397)	(0.00423)	(0.00526)	(0.00522)	(0.00473)	(0.00484)	(0.00507)	(0.00523)	(0.00505)	(0.00514)
Controls:											
Demographic controls		x	x	x	x	x	x	x	x	x	x
High school fixed effects				x	x	x	x	x	x	x	x
26 fields of study						x					
100 groups of degrees							x				
Course-university fixed effects								x		x	x
Maths content of course-university									x		
Family characteristics										x	x
Observations	53,851										
R-squared	0.028	0.230	0.240	0.365	0.378	0.443	0.457	0.502	0.387	0.502	0.504

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Notes: The sample consists of masters' and single-cycle college graduates from 2010-2012 cohorts interviewed both at graduation and 3 years later, born in Italy and residing in Italy upon graduation and who graduated between 23 and 31 years old, who are employed in full-time jobs 3 years after graduation and for whom there is info on all the variables of interest. Demographic controls include municipality of birth, year and age at graduation, province of job. High school controls include final grade and track in 8 categories (column (3)) or school-track dummies (column (4)). Academic controls include college final grade, GPA and time for completion, experiences abroad, language and IT skills, post-graduate training, and 26 fields of study (column (6)) or 100 classes of study (column (7)) or 3,300 university-degree courses dummies (column (8)) or maths content index for 3300 university-degree courses (column (9)). Family characteristics are the level of education of father and mother and the social class of the family (column (10)). Preferences are the variables taking value 1 if the student indicates as very important for the future job the following aspects: salary, career prospects, culture, stability, and free time for job (column (11)).

Table 1.7: Oaxaca Decomposition of the Gender Earnings Gap

	All workers		Full time sample	
Specification	(1)	(2)	(3)	(4)
Estimated gender gap	**-0.257*** **		**-0.146*** **	
	(0.00465)		**(0.00400)**	
Endowments:				
Overall	**-0.120*** **	**-0.127*** **	**-0.0699*** **	**-0.0645*** **
	(0.00538)	**(0.00520)**	**(0.00416)**	**(0.00371)**
Field of study	**-0.0396*** **	**-0.0380*** **	**-0.0535*** **	**-0.0397*** **
	(0.00350)	**(0.00194)**	**(0.00318)**	**(0.00189)**
Academic variables	-0.00271***	-0.00453***	0.000685	-0.000959
	(0.000985)	(0.000920)	(0.000790)	(0.000794)
Other skills	-0.00156	-0.00193	-0.00209	-0.00205
	(0.00158)	(0.00162)	(0.00156)	(0.00165)
Attitudes	**-0.0101*** **	**-0.0118*** **	**-0.00875*** **	**-0.0103*** **
	(0.00111)	**(0.00117)**	**(0.000971)**	**(0.00105)**
HS performance	0.000819	0.00304***	0.000839	0.00252***
	(0.000656)	(0.000669)	(0.000634)	(0.000644)
HS track	0.00655*	0.000623	-0.00686**	-0.0135***
	(0.00351)	(0.00369)	(0.00318)	(0.00339)
Family characteristics	9.75e-05	-0.000320	-0.000266	-0.000556
	(0.000415)	(0.000444)	(0.000387)	(0.000414)
Full time dummy	-0.0735***	-0.0745***		
	(0.00256)	(0.00270)		
Coefficients:				
Overall	**-0.137*** **	**-0.129*** **	**-0.0761*** **	**-0.0815*** **
	(0.00698)	**(0.00679)**	**(0.00502)**	**(0.00526)**
Field of study	0.0189	**-0.0349*** **	-0.0478	**-0.0281*** **
	(0.0732)	**(0.00366)**	(0.0786)	**(0.00410)**
Academic variables	-0.00224	-0.0562*	-0.0118	-0.0436
	(0.0298)	(0.0313)	(0.0313)	(0.0335)
Other skills	-0.0120	-0.0127	0.00416	0.00591
	(0.0313)	(0.0327)	(0.0347)	(0.0372)
Attitudes	0.00607	0.00760	0.00151	0.00228
	(0.00675)	(0.00743)	(0.00670)	(0.00742)
HS performance	0.00121	0.00306	0.00597	0.00708
	(0.00502)	(0.00509)	(0.00513)	(0.00502)
HS track	-0.00279	-0.00317	0.0106	0.0138
	(0.0118)	(0.0124)	(0.0116)	(0.0119)
Family characteristics	0.0169*	0.0152	0.0127	0.00661
	(0.00981)	(0.0105)	(0.00994)	(0.0105)
Full time dummy	0.0510***	0.0579***		
	(0.0103)	(0.0111)		
Constant	-0.214**	-0.106**	-0.0514	-0.0454
	(0.0868)	(0.0479)	(0.0920)	(0.0510)
Observations	71,220		53,851	

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Notes: Field of study is 100 classes in specifications (1) and (3) and maths content of degree course in specifications (2) and (4).

Appendices

Table A1: Sample Selection

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	All 2010-2012	Interviewed	Italian 23-32	Employed	Full time	Final sample	Final sample FT
Observations	218,505	154,777	141,828	93,006	70,333	71,220	53,851
VARIABLES							
Interviewed at graduation	0.91						
Interviewed three years after	0.71						
Female	0.60	0.61	0.61	0.58	0.52	0.59	0.53
Age	30.42	30.15	29.26	29.31	29.25	29.26	29.20
Employed		0.67	0.66				
Employed full time				0.76		0.76	
Average monthly wage						1154	1301
Missing wage				0.04	0.04		
Field of Study							
Education	0.02	0.02	0.02	0.02	0.01	0.02	0.01
Humanities&Arts	0.14	0.13	0.13	0.13	0.11	0.14	0.11
Social sciences, business and law	0.38	0.38	0.38	0.37	0.36	0.38	0.38
Science, Maths and Computing	0.08	0.08	0.09	0.07	0.07	0.08	0.08
Engineering, Manufacturing	0.20	0.21	0.22	0.27	0.32	0.24	0.29
Agriculture	0.02	0.02	0.02	0.03	0.03	0.03	0.03
Health and Welfare	0.15	0.14	0.13	0.10	0.09	0.10	0.09
Services	0.01	0.01	0.01	0.01	0.01	0.02	0.01
Academic variables:							
Graduation grade	107.1	107.3	107.5	107.2	107.0	107.3	107.2
Late degree (index)	0.26	0.25	0.24	0.25	0.25	0.25	0.24
GPA 18-26	0.38	0.37	0.36	0.38	0.39	0.36	0.37
GPA 27-29	0.45	0.46	0.46	0.46	0.46	0.47	0.47
GPA 29-30	0.17	0.18	0.18	0.16	0.15	0.18	0.16
Other training		0.72	0.74	0.66	0.64	0.67	0.65
Number of foreign languages:							
0		0.26	0.25	0.25	0.25	0.26	0.26
1		0.45	0.46	0.46	0.47	0.47	0.48
2		0.21	0.21	0.21	0.21	0.22	0.22
3		0.04	0.04	0.04	0.04	0.04	0.04
4		0.00	0.00	0.00	0.00	0.00	0.00
missing		0.04	0.04	0.03	0.03		
IT skills (number of IT tools):							
0		0.03	0.03	0.02	0.02	0.02	0.02
1-2		0.14	0.14	0.12	0.11	0.12	0.11
3-4		0.33	0.33	0.33	0.32	0.34	0.33
5 or more		0.49	0.50	0.52	0.55	0.52	0.54
Period abroad		0.20	0.21	0.22	0.23	0.22	0.23

continuing

Table A1 continued

	(1) All 2010-2012	(2) Interviewed	(3) Italian 23-32	(4) Employed	(5) Full time	(6) Final sample	(7) Final sample FT
High School:							
Final grade 60-84	0.44	0.43	0.42	0.44	0.43	0.43	0.43
Final grade 84-95	0.23	0.24	0.25	0.25	0.25	0.25	0.25
Final grade 95-100	0.30	0.32	0.34	0.31	0.32	0.32	0.32
Final grade missing	0.03	0.02	0.00	0.00	0.00		
Curriculum:							
Classics	0.19	0.19	0.19	0.16	0.14	0.16	0.15
Education	0.06	0.06	0.06	0.06	0.04	0.05	0.04
Languages	0.07	0.07	0.07	0.07	0.06	0.07	0.07
Art	0.02	0.01	0.01	0.02	0.01	0.01	0.01
Technical non-STEM	0.12	0.13	0.12	0.13	0.13	0.13	0.14
Technical STEM	0.08	0.09	0.08	0.10	0.12	0.10	0.11
Scientific	0.43	0.44	0.46	0.46	0.48	0.46	0.47
Professional	0.02	0.02	0.01	0.01	0.01	0.01	0.01
Foreign school	0.01	0.01	0.00	0.00	0.00		
Family characteristics:							
Father education:							
Less than high school		0.33	0.32	0.32	0.32	0.33	0.32
High School:		0.41	0.42	0.43	0.44	0.44	0.44
College degree		0.24	0.24	0.23	0.23	0.23	0.23
missing		0.02	0.02	0.02	0.02		
Mother education:							
Less than high school		0.33	0.31	0.32	0.31	0.32	0.32
High School:		0.44	0.46	0.47	0.47	0.48	0.48
College degree		0.21	0.21	0.20	0.20	0.20	0.20
missing		0.02	0.02	0.02	0.02		
Social class:							
Managerial and professional workers		0.25	0.26	0.25	0.26	0.25	0.26
Intermediate occupations		0.31	0.32	0.31	0.32	0.32	0.32
Non professional self-employed		0.20	0.20	0.21	0.21	0.21	0.21
Routine occupations		0.22	0.22	0.22	0.21	0.22	0.21
missing		0.02	0.02	0.02	0.02		
Preferences							
Importance salary		0.54	0.54	0.54	0.55	0.54	0.55
Importance career prospects		0.62	0.62	0.63	0.65	0.63	0.65
Importance job stability		0.65	0.65	0.64	0.63	0.65	0.64
Importance culture		0.45	0.45	0.42	0.39	0.42	0.40
Importance free time		0.24	0.23	0.22	0.21	0.22	0.21

Notes: Summary statistics for: the overall number of masters' and single-cycle college graduates from 2010-2012 cohorts (column (1)), who are interviewed both at graduation and 3 years later (column (2)), born in Italy and residing in Italy upon graduation and who graduated between 23 and 31 years old (column (3)), employed three years after graduation and with all non-missing observations (respectively columns (4) and (6) for all workers and columns (5) and (7) for full time workers only).

Table A2: Summary Statistics by Employment Status and Gender

	Males		Females	
	Employed	Unemployed	Employed	Unemployed
	(1)	(2)	(3)	(4)
VARIABLES				
Age	29.44	29.59	29.21	29.33
Employed full time	0.86		0.68	
Missing wage	0.04		0.03	
Field of Study				
Education	0.00	0.00	0.03	0.03
Humanities&Arts	0.06	0.12	0.18	0.18
Social sciences, business and law	0.31	0.45	0.41	0.52
Science, Maths and Computing	0.08	0.10	0.07	0.09
Engineering, Manufacturing	0.42	0.21	0.15	0.09
Agriculture	0.03	0.03	0.02	0.02
Health and Welfare	0.08	0.06	0.12	0.06
Services	0.02	0.02	0.01	0.01
Academic variables				
Graduation grade	106.2	106.0	107.8	107.6
Late degree (index)	0.27	0.27	0.24	0.25
GPA 18-26	0.45	0.44	0.32	0.34
GPA 27-29	0.43	0.40	0.48	0.47
GPA 29-30	0.12	0.15	0.19	0.19
Other training	0.61	0.84	0.70	0.84
Number of foreign languages:				
0	0.26	0.27	0.25	0.26
1	0.53	0.47	0.41	0.41
2	0.16	0.19	0.25	0.23
3	0.02	0.03	0.06	0.05
4	0.00	0.00	0.00	0.00
missing	0.03	0.04	0.04	0.05
IT skills (number of IT tools):				
0	0.02	0.03	0.03	0.03
1-2	0.08	0.13	0.15	0.17
3-4	0.27	0.29	0.37	0.34
5 or more	0.63	0.55	0.44	0.45
Period abroad	0.21	0.19	0.22	0.16
Preferences				
Importance salary	0.55	0.56	0.53	0.59
Importance career prospects	0.68	0.66	0.59	0.64
Importance job stability	0.58	0.64	0.69	0.74
Importance culture	0.36	0.43	0.46	0.50
Importance free time	0.21	0.26	0.22	0.25

	Males		Females	
	Employed	Unemployed	Employed	Unemployed
	(1)	(2)	(3)	(4)
High School:				
Final grade 60-84	0.51	0.54	0.39	0.40
Final grade 84-95	0.23	0.23	0.26	0.27
Final grade 95-100	0.26	0.23	0.35	0.34
Final grade missing	0.00	0.00	0.00	0.00
Curriculum:				
Classics	0.11	0.18	0.20	0.26
Education	0.01	0.01	0.09	0.11
Languages	0.02	0.02	0.11	0.09
Art	0.01	0.01	0.02	0.02
Technical non-STEM	0.11	0.13	0.14	0.13
Technical STEM	0.21	0.15	0.02	0.02
Scientific	0.53	0.48	0.40	0.36
Professional	0.01	0.02	0.01	0.01
Father education:				
Less than high school	0.29	0.30	0.35	0.36
High School:	0.43	0.41	0.43	0.42
College degree	0.26	0.26	0.21	0.19
missing	0.02	0.02	0.02	0.03
Mother education:				
Less than high school	0.30	0.31	0.34	0.35
High School:	0.47	0.44	0.46	0.44
College degree	0.21	0.23	0.18	0.18
missing	0.02	0.03	0.02	0.02
Social class:				
Managerial and professional workers	0.27	0.25	0.23	0.20
Intermediate occupations	0.32	0.33	0.31	0.32
Small Employers and non professional self-employed	0.19	0.17	0.22	0.20
Routine occupations	0.21	0.24	0.22	0.26
missing	0.01	0.02	0.02	0.03
Observations	39,141	8,015	53,865	18,538

Notes: Summary statistics for masters' and single-cycle college graduates from 2010-2012 cohorts, by employment status and gender.

Chapter 2

Early Influences and the Gender Gap in STEM

2.1 Introduction

During the past 40 years there has been a striking reversal of the gender gap in education in industrialised countries. Although women are currently more likely than men to hold a college degree in the vast majority of OECD countries, their choices of college major have been and persistently continue to be different from those of men. Figure 2.1 illustrates the percentage of females among graduates with a bachelor degree in 7 OECD countries in 2015, for all fields of education and separately for the fields of science, engineering, education and humanities. In all countries but Germany women constitute more than half of all bachelor's degree graduates and are greatly over-represented in education and humanities, but they represent only 20 to 30% of engineering graduates.

Science, technology, engineering and mathematics degrees – indicated with the acronym STEM – have been the object of increasing attention in education, economic and policy fora. During the 2017 celebration of the International Day of Women and Girls, the UN Assistant Secretary-General Lakshmi Puri stated that *"we must ensure that women's participation in innovation is not the exception, but becomes the norm"*. Several initiatives aimed at encouraging female students to undertake STEM careers have been promoted all around the world; some examples are the initiative '*Girls in*

Stem' in Turkey from the Nobel Laureate in chemistry Professor Aziz Sancar and the '*Girls in ICT*' from the International Telecommunication Union. In Italy, which is the setting for the present study, the Gender Equality Department of the government launched '*Stem Month*' in 2016, showcasing a series of initiatives targeting female pupils in primary and secondary schools, with the goal of encouraging their interest in STEM subjects.

There is a widespread consensus that STEM skills are crucial to sustaining innovation and growth (Osikominu et al., 2014). However, the share of graduates in STEM majors across OECD countries in 2015 was only 23% (and the enrolment share was approximately 27%). Thus, understanding the mechanisms underlying the educational segregation of women may shed light on issues regarding the scarcity of scientists that the European Union is concerned about.

Furthermore, several studies have provided evidence that – because STEM degrees typically lead to higher-paying jobs – gender gaps in college majors translate into gender gaps in earnings later in life (Flabbi, 2012; Anelli and Peri, 2015a; Card and Payne, 2017).

In this paper, I analyse the determinants of gender gaps in STEM graduation rates for Italian college-leaving cohorts from 2010 to 2015, with an emphasis on family, cultural and school influences, as well as geographic proximity in the supply of STEM degrees. For this purpose I use data from a uniquely rich and largely unexplored source (AlmaLaurea) that combines both administrative and survey information on the population of Italian graduates.

I am able to characterise the students' pre-college education in its most relevant aspects. One aspect is the curriculum of the high school attended, which varies widely in its maths components across a large number of available tracks. Moreover, a secondary school identifier allows me to capture the influence of unobservable school characteristics, over and above differences in their official curriculum. These administrative data are supplemented by survey-based information on students' family background and their attitudes and aspirations. By exploring the role of gender preferences in shaping college major choices I contribute to the literature on the impact of gender differences in personal traits – largely documented by the experimental literature[1]– on real-life choices.

I complement the data from AlmaLaurea with information on the general attitudes, demographic composition and political orientation of Italian municipalities. This information is then used to characterise the elements of students' background that are arguably related to gender identity norms. Finally, I use administrative data on the supply of STEM degree programmes across Italian universities in order to relate students' choices of majors to the geographic distribution of the supply of STEM degrees.

I estimate an average unadjusted gender gap in STEM graduation rates of approximately 22 percentage points for 2010-2015 cohorts. The most important determinant of this difference, driving approximately half of the observed gap, is the gender difference in the maths and science content of the respective high school curricula. This difference can be traced to educational choices made at age 14, when boys are more likely than girls to enrol into high school tracks that are more intensive in maths and science. Despite differences in high school choices, girls on average complete high school with a higher final grade than boys, regardless of track. This result implies that if girls were under-performing relative to boys in maths- and science-intensive high school tracks, the gender gap in major choices would be even greater. Based on self-reported measures of students' personal traits, the attitudes of girls suggest lower competitiveness and higher altruism and social mindedness; however, these differences do not appear to play an important role in driving the gender difference in major choices. On the other hand, male and female students have, on average, very similar family and social environments – as measured by the parental and municipality characteristics. Therefore, the gender gap in the outcome cannot be explained by differences in these environments.

When this large set of characteristics is controlled for, half of the gap remains unexplained. The results from an Oaxaca decomposition show that approximately 50% of the part of the gap not explained by difference in characteristics is accounted for by a much lower probability of girls of choosing a STEM degree even conditional on having attended one of the maths- and science-intensive high school tracks. The

[1]See Azmat and Petrongolo (2014) for a review of this literature.

results also suggest that family and social background features – over and above the influence they can already have on attitudes and previous choices – affect female and male college choices differently, each accounting for another 20% of the unexplained part of the STEM gap.

The remainder of the chapter is organised as follows. Section 2.2 describes the conceptual framework and reviews the related literature; Section 2.3 describes the background of STEM college majors in the Italian education system. A description of the data and summary statistics are provided in Section 2.4. Section 2.5 presents and discusses the results based on the Gelbach and Oaxaca decompositions of the estimated gender gap in the choice of a STEM major. Section 2.6 concludes the chapter.

2.2 The determinants of major choice

In this section I discuss the factors and mechanisms potentially shaping the gender gap in major choices in greater detail. I focus on three sets of explanations: (i) human capital factors, i.e., a student's preparation and achievement at pre-collegiate levels of education; (ii) personal factors, summarised by individuals' attitudes and aspirations for their future career; and (iii) parental and societal influence, which can in turn affect both high school choices and individuals' preferences for higher education.

2.2.1 Pre-college education

The choice of enrolling in a STEM university course is realistically influenced by the science and maths ability and knowledge that students would have acquired prior to choosing their major. These ability and knowledge are in turn largely determined by the high school track attended. In Italy, the first stage of education that offers a range of curricular choices is the start of high school, which follows the completion of middle school at age 14. Tracks available may be academic or vocational, and they vary widely in maths content. Within the academic system, high schools ("licei") specialise in one of the following: maths and science, humanities, modern languages or art. Within the vocational system, high schools ("istituti") offer a wide variety of tracks with specialisations in IT and technical applications, business and accounting, administration, tourism, etc. The distinction between the academic and vocational tracks was originally conceived to prepare students for higher education and middle-skill-level jobs, respectively. Following a law approved in 1969[2], students graduating from any high school have access to higher education. An important point to note is that in the Italian education system the choice of curriculum is made at the relatively early age of 14, when family influences may be stronger than they are later in life.

The existing literature has investigated whether boys and girls make systematically different choices prior to college entry. For the US, Xie and Schauman (2003) find that girls are less likely than boys to participate in science and engineering courses in high school. For Canada, Card and Payne (2017) find instead that the gender gap in the fraction of high school graduates who have taken STEM courses is small and is not the main explanation for the gender gap in STEM majors. My evidence for Italy demonstrates that girls are largely under-represented in maths-intensive high school tracks. In my final sample of college graduates, only 53% of girls have completed maths-intensive or technical high schools, in contrast to 83% of boys. The extent to which this gap maps to gender gaps in college majors depends on the explanatory power of the high school track in shaping major choices. Evidence for both the US and the UK indicates that taking maths-intensive courses in high school is a strong predictor of a later STEM major choice (Gottfried and Bozick, 2016; Philippis, 2017).

Secondary education may also impact major choices via specific (observable or unobservable) high school characteristics, over and above their general track. For example, Legewie and DiPrete (2014) find that, all else being equal, gender segregation in extra-curricular activities have a discernible impact on the gender gap in the STEM choice in US. This evidence may be consistent with the self-selection of girls into high school with certain characteristics predictive of STEM choice, or with a differential gender impact of such characteristics.

Finally, conditional on high school choice, performance and final grades may play a role in STEM choice. STEM degrees are typically considered the most demanding

[2] Law n.910 of the 11th of December 1969.

ones; in a sample of higher education graduates from 14 OECD countries, Flabbi (2012) finds that science fields attracts the highest proportion of top-performing students in secondary school in both the male and female samples. Moreover, when looking at the perceived characteristics of the study programme, he finds that more than 20% of men and women regard study programmes in the scientific field as very demanding, while only approximately 10% of the respondents express the same judgement about humanities programmes of study. I find evidence that better high school grades are positively associated with later pursuing a STEM degree; this observation is interesting given that girls in my sample achieve, on average, better high school final grades than boys regardless of track.

2.2.2 Personality traits

Preferences are arguably an important factor in major choice. Wiswall and Zafar (2015) observed that the single largest factor in determining a student's college major is represented by preferences and tastes – i.e., how much the individual likes the subject and the job associated with it. This is even after randomly providing some students with additional information, such as earnings potential associated with the different majors.

Several recent studies have demonstrated that men and women are systematically different in some psychological attributes.[3] Females are found to be more risk averse and less willing to compete, and this could explain why they choose careers with less risk and competition. Moreover, women are found to be more socially minded and altruistic, which may translate into different occupational aspirations and career preferences. Such differences could be associated with differences in major choices, as majors in humanities and social sciences may be associated with a larger interest in society, while maths-intensive majors such as engineering may be associated with a more egoistic and competitive view of the world (Anelli and Peri, 2015b).

The evidence on the influence of these differences on real-life choices is not very rich and is mainly constrained by the lack of data adequately measuring personal

[3]See, for example, Booth and Nolen (2009), Gneezy et al. (2003), Niederle et al. (2013), Andreoni and Vesterlund (2001), Eckel and Grossman (1998).

traits. With respect to gender differences in college major choices, Zafar (2013) attributes the gender gap mostly to gender differences in preferences and tastes, particularly to men's stronger emphasis on pecuniary outcomes and women's stronger emphasis on enjoying their coursework and employment in potential jobs. My evidence is consistent with the following assumptions for females (compared to males): earnings are less important while culture is more important; career prospects count less, suggesting lower competitiveness; free time is valued more; and women are more involved in volunteering activities, which suggests greater social mindedness and altruism.

2.2.3 Family and social background

The seminal work of Akerlof and Kranton (2002) introduced the idea that individuals' social identity enters into their choices, and thus social incentives may explain why observed choices are at odds with economic incentives. Applying this idea to the gender gap in major choice implies that certain women with high ability may choose to exert lower effort and select less difficult majors with lower monetary returns when identity enters their choices, because it is expected from them under the prevailing gender identity norms and they internalise social expectations about their role. External influence can originate from a close environment, such as the family, or from broader social settings in which individuals live, such as the civic community.

A vast body of literature demonstrates positive correlations between parents and children in terms of economic, educational, social, and behavioural outcomes. Parents' educational achievement is important to the extent that it proxies parents' abilities and skills, which are strong predictors of the abilities and skills of their children.[4] Several studies emphasise that the family environment is relevant for the transmission not only of skills but also of gender norms, and they document a positive correlation between the gender role attitudes of parents and children.[5] Cheng et al. (2017) provide interesting evidence of maternal role modelling for daughters'

[4]For an extensive review of the literature on the intergenerational transmission of education and earnings see Black and Devereux (2011).

[5]For example, Farré and Vella (2012) find that in a sample of US mothers and children, children's views about working women are affected by their mother's attitudes, which in turn influence female labour market decisions.

choices: they find that having the mother employed in a STEM occupation increases the probability of the child working in hard sciences. Thus, measuring aspects of the family arguably related to attitudes towards females, including the education or employment/social status of the mother relative to that of the father, is important in studies focusing on young students' choices.

In addition, the civic community in which individuals grow up can be important for the transmission of gender norms. Several studies indicate a direct relationship between attitudes towards women and the maths gender gap in a given society. For example, Guiso et al. (2008) compare gender differences in test performance across countries with different levels of gender equality and find that girls' under-performance in maths relative to boys' performance is eliminated in more gender-equal cultures. Moreover, González de San Román and de la Rica (2012) find that girls perform relatively better in both maths and reading in societies where gender equality is enhanced, and Nollenberger et al. (2016) demonstrate that the maths gender gap for each immigrant group living in a particular host country (and exposed to the same host country's laws and institutions) is explained by measures of gender equality in the parents' country of ancestry. The influence of the social environment can be particularly relevant in a context such as Italy, where there is a high degree of cultural diversity even across small communities such as the municipalities.

2.3 STEM in the Italian context

The acronym STEM refers to a *"group of disciplines that teach the skills required for a high-tech economy"*.[6] What this means in practice, as well as how this definition relates to specific courses in higher education institutions, is a more complex matter; the definition varies across countries, and sometimes even among different bodies within the same country.

In Italy, a list of the university courses that are considered STEM is provided by the Ministry of Education (MIUR). These are the courses that correspond to groups 04 and 05 of the classification FOET (Fields of Education and Training)

[6] Definition from the House of Lords 2nd Report 2012-2013 on Higher Education in STEM subjects.

1999: 'science, mathematics and computing' and 'engineering, manufacturing and construction'[7]. In table 2.1 I report the FOET 1999 classification in terms of both broad fields and a finer classification based on 'fields of education'. Within the two STEM groups, we can distinguish 7 fields: life sciences, physical sciences, maths & stats, engineering, manufacturing, architecture and building, and computing.

The STEM definition appears to include a fairly heterogeneous group of fields of study. I look at administrative data on students' enrolment in Italian universities in 2010 – made available by the MIUR – to analyse the gender gap in enrolment by field of study. The overall gender gap in enrolment in STEM fields in 2010 was 19 percentage points, with the average probability of enrolling in a STEM degree being 27%. When analysing the enrolment gender gap for each of the sub-fields (Figure 2.2), I find a relevant degree of heterogeneity.[8] Within STEM fields (panel (a)), the gender gap is more pronounced in some fields including computing and engineering, physics and earth science. By contrast, for other fields such as architecture, chemistry, and maths & stats the gap is smaller, or even reversed, as for manufacturing and life sciences. On the other hand, most non-STEM fields (panel (b)) are characterised by a positive gender gap; the exceptions are business and administration and most of the service fields.

To identify the characteristics that distinguish fields in which females are more likely to enrol from fields that are male-dominated, I use administrative data from the MIUR on the very detailed content of each of the approximately 2,500 unique undergraduate or single-cycle courses offered by Italian higher education institutions in 2010. I characterise the maths content of each course by building a *maths intensity* index, which is the proportion of university 'credits' that students have to obtain in maths-intensive subjects out of all the credits they need in order to graduate from a specific course. Across all courses classified as STEM, the average index is 0.64, while for non-STEM courses it is 0.13: STEM courses are clearly the maths-intensive ones. Figure 2.3, which plots the index separately for each STEM and non-STEM

[7] Geography is classified as physical science and is in group 04, but it is excluded from the STEM definition.

[8] I adopt here a further classification for the physical sciences group – namely, distinguishing physics, chemistry, and earth sciences – and for the architecture and building field – distinguishing architecture and town planning from building and civil engineering.

sub-field, respectively in panel (a) and panel (b), shows that maths intensity varies substantially across different fields. Within STEM fields, life science, chemistry and earth science are characterised by a relatively low maths content. Within non-STEM fields, business and administration, transport service and security service fields are characterised by a relatively high maths content.

The analysis of course content and of enrolment patterns points to a negative correlation between the maths intensity of a field and the gender gap in the probability of enrolling in majors in that field. Figure 2.4 plots the maths intensity and enrolment gender gap of the different fields of study on the x-axis and the y-axis, respectively. The majority of the STEM fields fall in the bottom right part of the graph; i.e., they are characterised by high maths content and a negative gender gap in enrolment. The opposite is true for most non-STEM fields. Within STEM fields, the ones characterised by a relatively lower gender gap in enrolment are also the ones with less maths content (for example chemistry, earth and life sciences), and the opposite is true within non-STEM fields (for example, business and administration and most of service fields). The correlation between these two measures is -60%. Even at the level of more than 2,000 unique university courses, the correlation is almost -50%.

I will use the information obtained on course content to estimate the gender gap in the maths intensity of the specific course of study chosen and analyse its determinants.

2.4 Data and Variable Description

To analyse students' choices of major, I use data from the *AlmaLaurea Graduates' Profile*, a survey of the population of college graduates from most Italian universities interviewed upon graduation, which is made available by the research institution AlmaLaurea. I focus on students from undergraduate and single-cycle courses graduating from 2010 to 2015 in one of the 56 universities taking part in the survey for the whole period considered.

Not all students enrolled in universities will obtain a degree, and in this sense, the AlmaLaurea database represents only a selected sample of students. In particular, if the drop-out rate is differential between male and female students, this might result in an over- or underestimation of the real gender gap in the choice of studying a STEM subject. The direction of the bias is not clear *a priori*: female students might be more likely to be discouraged than male students because of their different attitudes towards competition, or women may be influenced by social pressures based on the belief that they are less suitable than men for such careers and may thus be more likely to drop out. It is also possible that only the most determined females enrol in STEM, such that STEM female students are less likely than males to drop out.

Enrolment data are available from the MIUR for the years since 2003, only aggregated at the university, field of study and province of residence level. I compare the graduation rates obtained from the AlmaLaurea data with data on enrolment rates in STEM fields by gender and year of enrolment. Figure 2.5 is a plot the obtained graduation and enrolment rates and the gender gaps. The graph illustrates the lack of association between the drop-out rate in STEM fields and gender, indicating that the gender gap in graduation is a good proxy for the gender gap in the choices made by young students at time of enrolment. Given that the outcome analysed in this study is a rate resulting from the joint probability of enrolling in a STEM degree and of graduating with a STEM degree, the results of the analysis should be interpreted while noting that the impact of any factor on this outcome entails both the impact on the decision at the time of enrolment and the impact on subsequent decisions up to graduation.

For the purpose of my analysis, I exploit the richness of the *Graduates' Profile* survey to gain access to several pieces of information about each student's background. I am particularly interested in three groups of variables: (i) graduates' high school choices and performance, (ii) their attitudes and aspirations, (iii) their family and social background.

Administrative variables provided by each university include: high school final grade; high school curriculum, which gives a useful measure both of students' preferences at earlier stages in life and of the type of skills they have at the moment of enrolling in the university; and names of the specific high schools attended by each student, which allows to control for the role of other high school characteristics over and above their general track.

The other variables are constructed from students' answers to the questionnaire. I measure students' attitudes and aspirations through answers to questions on: the motivation for the major choice, particularly whether professional or cultural factors had a greater influence on the decision; the relevance of several aspects related to their future career, including salary, career prospects, culture, stability and free time; the engagement in volunteering activities, which can be regarded as reflecting how altruistic and socially minded an individual is.

To characterise a student family background, I draw on answers to questions about the level of education of both parents and their last occupation to proxy for socio-economic status. An interesting aspect of the survey is that it collects information on the field of study for parents with college degree. This information helps to distinguish and evaluate the importance of whether the students' mother and father have a STEM degree relative to other degrees.

2.4.1 Local variables from other data sources

An important piece of information for my analysis in the AlmaLaurea survey is the municipality of origin of each graduate. Universities provide both the municipality of birth and the municipality of residence at the time of enrolment. I draw on the latter to characterise a student's sociocultural background at the time of major choice. Secondary data sources are used to construct alternative indicators for society progressivism at different time periods and in different municipalities. The goal is to recover some indirect measures of women equality in Italian society along two different dimensions: political empowerment and sexual emancipation.

To measure women's political empowerment, I use an indicator of whether the mayor is a female and the share of females in municipal councils, both taken from the *Census of Local and Regional Administrators* made available by the Italian Ministry of the Interior.

Following Braga and Checchi (2008), I use as proxies for women's sexual emancipation the municipality-specific fertility rate – calculated as the number of live births divided by the number of women between ages 15 and 49 times 1,000 – and the share of religious marriages over the total number of marriages, both obtained from the "*Atlante Statistico dei comuni*" of the Italian National Institute of Statistics (ISTAT). As women's control over their sexuality increases, the fertility rate should decrease. Civil marriages are characterised by lower gender segregation and a greater equality between partners.

I am able to build a consistent time series for the period between 2003 and 2011. In Figure 2.6, I plot the variables for 2010. Only 10% of the municipalities are governed by a female mayor, and panel (a) of the figure illustrates that these municipalities are concentrated in the northern part of the country. On average across all municipalities, the share of female councillors in local governments is only 20%, and as depicted in panel (b) the percentage is higher in northern municipalities. The average fertility rate is approximately 39 across all municipalities, and panel (d) shows that fertility is unexpectedly higher in northern regions than in southern regions, although the geographical pattern is not very clear and sharp. Finally, most marriages in Italy are celebrated with religious rituals: on average, the percentage of total marriages is 68%, and as shown in panel (d), the rate is higher in southern Italy.

2.4.2 Supply of STEM education

Students' decision to enrol in a STEM degree programme is potentially also a function of the availability of STEM courses. A student residing in a given municipality upon finishing high school faces a distribution of university courses offered in different locations across the country. The student's choice of major then depends not only on his/her preferences but also on the characteristics of this supply.

I use administrative data on higher education made available by the MIUR to measure the different factors characterising the higher education supply in Italy, and I summarise them in a single *supply index*. In particular, for each STEM and non-STEM course available, I extract the geographical location in which it is offered, the size of the university offering it and the availability of scholarships at the university.

An Italian student with a general high school degree can in principle choose from all of the available tertiary education programmes and institutions. For a specific group of majors – namely, most majors in the health group (medicine, dentistry) plus architecture and the recently established (2008) major educational science – access is limited and conditional on the successful performance on entry tests, which are managed nationally by the MIUR. For other majors, each offering institution can decide to set a limit on the number of students who can enrol each year. Unfortunately, information on the exact number of places made available by each university for each major characterised by nationally or locally managed limited access is not available. This makes it impossible to construct a precise measure of the availability of places supplied by each university for every field of study. By contrast, data on the number of students enrolled yearly in each major at different universities, which are easily accessible, give a measure of the equilibrium quantity resulting from the supply and demand for education. At best, this measure can be used as a proxy for the quantity of supply. In particular, I use data on enrolment to classify universities into 4 categories: very large (more than 40000 students enrolled), large (between 20000 and 40000 students enrolled), medium (between 10000 and 20000 students enrolled), and small (less than 10000 students enrolled).

The enrolment choice is also constrained by costs. Direct pecuniary costs depend on tuition fees and scholarships availability. In Italy, tuition fees are relatively low compared to international equivalent, they are similar across universities (except for a few private ones) and vary insignificantly across majors within a university. On the other hand, the availability of scholarships can vary substantially among different institutions: the level of scholarships awarded to eligible students depends on the availability of regional funds, that can vary greatly among regions. Typically, southern regions are characterised by lower availability of regional funds and consequently of scholarships relative to those available in northern regions. I draw on data on the percentage of scholarships awarded to eligible students to construct weights that confer higher relevance to universities in which the likelihood of receiving a scholarship is higher.

Another important aspect of the cost of choosing a given course of study is represented by the geographical proximity to the municipality where the course is offered. I calculate the linear distance from each Italian municipality to each municipality where a higher education course is offered. Based on the calculated distance, I construct a geographical proximity weight. This value is always 1 if the linear distance is 0 (the course is offered in the same municipality); for other municipalities, it is the inverse of the linear distance.

For each Italian municipality I construct an index by summing the number of courses – both overall and of STEM fields only – offered in all Italian municipalities, weighted by the following: the size of the university offering the course, the percentage of scholarships awarded to eligible students at each university, and the geographical proximity to the municipality where the course is offered.

Figure 2.7 is a plot of the resulting 2010 index for the overall supply and the STEM supply by municipality. The supply of STEM education is clearly correlated with the overall supply, but not perfectly. The figures show the dramatic difference in the supply of higher education between northern and southern Italy. Students residing in northern Italy clearly face a higher supply relative than do students coming from southern regions, and this variation may account for differences in STEM graduation rates between students from different parts of the country. Assuming that male and female students are equally distributed across municipalities, these differences in the supply measure should be less relevant for the gender gap. However, if female and male students respond differently to supply, then this variable might account for part of the gender gap. For example, females might be less likely than males to leave the family and move – because of different preferences or social attitudes towards females' choices. This would imply that, given the same distance from a STEM course, females may be less likely to enrol in such a course.

2.4.3 Final Sample and Summary Statistics

The number of college graduates from 2010-2015 cohorts exiting from one of the 56 universities taking part in the AlmaLaurea survey for the entire period considered is approximately 1.1 million.

To analyse the choice of field of study, I focus on 3-year undergraduate or 5-year single cycle students, numbering approximately 790,000. I restrict the sample to students who were born in Italy and residing in Italy at graduation – excluding

4% of the observations – and who enrolled between the ages of 18 and 21 years old in the years from 2003 to 2011 – approximately 80% of the sample – which are the years for which I have data on the variables at the municipal level.

I merge these data with the data on municipality characteristics and the local supply of STEM programmes. For approximately 85% of the observations I have information on all the variables, so the final sample consists of 485,350 observations.

Table 2.2 lists summary statistics of the main variables presented separately for male and female students in the sample. Females constitute 62% of the sample, confirming that women are over-represented in the population of university graduates. As expected, the outcome variable documents a large gender gap in the probability of graduating in STEM fields, precisely 22 percentage points, which is 85% of the overall average probability of studying STEM. When looking at maths intensity of the course chosen, I find a gender gap that is similar in magnitude: the percentage of maths-intensive subjects in courses chosen by females is, on average, 22 percentage points less than that for their male peers.

The distribution of the two samples across high school study paths shows that young girls are over-represented in the humanities track while boys mainly choose the scientific path.[9] The majority of men are tracked early on into classes with higher exposure to science and maths, and vice versa for girls. On the other hand, females always outperform males: they obtain a higher final high school grade on average regardless of the track chosen.

In terms of attitudes and aspirations, some interesting differences emerge: relative to men, women are less likely to declare that they have chosen their field of study for professional rather than cultural motivations, they are less likely to consider career prospects to be very important for their future job, and they seem to more strongly value aspects such as culture and stability of the job. Moreover, on average, female students carry out more volunteering activities than their male peers.

Furthermore, compared with males, females appear to have parents who are

[9]The Scientific & Technical category is an indicator for having attended a 'scientific' high school offering students a maths- and science-intensive curriculum or a 'technical' high school offering specialisation in technological subjects such as IT, electronics or chemistry. The Humanities category is an indicator for having attended humanities-intensive high schools including 'classics', 'languages' and 'artistic' tracks.

slightly less educated and have lower-level jobs.

The final group of variables included in the analysis are those measured in the municipality of residence in the year of enrolment at university, which are used to characterise the social background in which a student made the choice of major upon exiting from high school and the supply of higher education faced. Unsurprisingly, there is no difference between females and males in these variables. Thus, if any of these variable explains the gender gap in STEM graduation rates, this would not be due to differences in those environments but instead would stem from how the two sexes respond differently to similar environmental features.

2.5 Empirical Method and Results

I estimate a linear probability model for STEM major choice that takes into account human capital and personal factors, as well as family and societal influences. The specification estimated is given by:

$$y_{im\tau t} = \beta_1 F_i + X_i \beta_2 + Z_{m\tau} \beta_3 + \gamma m + \delta \tau + \eta t + u_{im\tau t} \qquad [2.1]$$

where $y_{im\tau t}$ is an indicator for graduation in a STEM field for student i who resides, upon enrolment, in municipality m, enrols in year τ and graduates in year t; F_i is a female dummy; X_i is a vector of individual and family characteristics; and $Z_{m\tau}$ is a vector of variables measured at the municipal level at the time of college enrolment. I also estimate the same specification for the outcome of the maths intensity index for the college course of study chosen by each student.

The results from the full regression estimations are reported in tables A1 and A2 of the appendix for the probability of graduating from a STEM major and for the maths intensity of the specific course attended, respectively. The results are very similar for the two outcomes. From the estimations performed on the pooled sample of females and males (columns (1) of both tables) we observe that having attended a maths- and science-intensive high school and having obtained a higher high school final grade are positively associated with both outcomes. Measures of personal traits that are arguably related to a higher level of competitiveness – such as professional rather than cultural motivation for major choice and the high value attributed to career prospects and salary for one's future job – are positively associated with the outcomes. On the other hand, personal traits suggesting lower competitiveness and higher social mindedness and altruism – such as the high value attached to culture and free time in one's future job and the participation in volunteering activities – are negatively related to the outcomes. A higher social status and a higher level of education of the two parents are associated both with a higher probability of graduating from a STEM major and with greater maths content of the college course. The association is stronger for parents with a STEM college degree and stronger for the father than for the mother. None of variables measured at the municipality of residence upon enrolment is significant in predicting the outcomes.

Given the estimate of the gender gap in the outcome $\hat{\beta}_1$, in order to identify and discuss the contributions of each of the five groups of variables – pre-college education, personal traits, family characteristics, social background and the supply of higher education – I adopt the conditional decomposition suggested by Gelbach (2016). Given the equation of the base model:

$$y_{im\tau t} = \tilde{\beta}_0 + \tilde{\beta}_1 F_i + \epsilon_{im\tau t} \qquad [2.2]$$

which gives the gender gap that we intend to decompose, Gelbach suggests a decomposition of the difference between the coefficients in the base model and the coefficient in the full model of equation [2.1], $(\hat{\tilde{\beta}}_1 - \hat{\beta}_1)$, given by the omitted variable bias formula: the difference is expressed as the product of the coefficient of each covariate in the full regression and the coefficient of a regression of the covariate on the female dummy. Thus, for each variable, we obtain a parameter measuring its contribution in explaining the gender gap, which is the female-male gap in the variable scaled by its STEM graduation/maths-intensity equation impact. Whether variation in a variable increases or reduces the gap depends on whether the covariate has a positive effect on the outcome and on whether the covariate has a higher mean for females or for males, thus the Gelbach decomposition gives a very useful and intuitive way of interpreting the contribution of each covariate in explaining the gender gap.

Table 2.3 reports the results from this decomposition of the coefficients both of the gender gap in STEM graduation and of the maths intensity of the university course. In columns (1) and (4) – respectively for the two outcomes – I report results from the estimation of a model where the high school curriculum is included in two categories: technical or scientific versus humanities. The high school track here explains approximately 18% of both outcomes. Among the other variables, differences in attitudes and in family characteristics each account for 2 to 4% of the gender gaps, while all the remaining variables together account for less than 1%.

In columns (2) and (5), I present results from a model in which I adopt a finer classification of the high school curriculum, which is the variable with the highest explanatory power. Within the humanities track, we can distinguish paths with a focus on classics, foreign languages, education or art; within the technical path, we can distinguish a group of tracks with a focus on business, tourism or agriculture (non-STEM) and another with a focus on industrial construction and preparation for surveyors (STEM). When the indicators for the 8 different high school tracks are included, this group of variables explains almost half (48%) of the gender gap in STEM graduation and almost 1/3 of the gap in maths intensity, while the role played by other groups of variables remains stable.

Next, I exploit the very detailed information on the secondary education institution attended by each student. I can distinguish approximately 5,500 different high schools attended by students in my sample. Some Italian high schools offer only one curriculum, while other larger ones can offer many different paths; thus, in the end, I have more than 11,000 school-track interactions. By including this information in my model, I am able to analyse the major choices conditioning not only on having chosen the same high school track but also on having attended the same secondary education institution. The results are presented in columns (3) and (6). Including the full set of school-track dummies leaves the results almost unchanged; thus, very little is due to differences in characteristics of schools attended by females and males other than their official curriculum.

The results from the Gelbach decomposition of the estimated gap in major choices indicate overall that, among the observable measured characteristics, the most important determinant of the gap is the gender difference in the maths and science content of students' high school curriculum. At the age of 14, boys and girls are already making different educational choices, with boys more likely than girls to enrol in high school tracks that are more intensive in maths and science. Differences in self-reported measures of students' personal traits do not appear to play an important role in driving the gender difference in major choice. As expected, since male and female students come, on average, from very similar family and social environments, differences in those environments fail to explain the gender gap in outcomes. Approximately half of the gap remains unexplained by differences in observed measured characteristics.

2.5.1 Oaxaca Decomposition

The analysis based on the estimation of model [2.1] assumes that the coefficients of the covariates are the same for females and males. To account for the difference in returns to the various characteristics, I perform an Oaxaca decomposition of the regression results from the estimation of the model that includes the high school track in 8 categories. The male-female difference in the outcome is decomposed in a portion that is 'explained' by group differences in characteristics and the residual portion that cannot be accounted for by such differences in the determinants of the outcome. The decomposition method is implemented such that the difference in characteristics is weighted by coefficients for males, while the difference in coefficients is weighted by characteristics of females.

The results for both outcomes are presented in table 2.4; all the predictors included in the regressions are summarised in five groups, as done above. The overall gender gap is explained in approximately the same proportion by the difference in coefficients and the difference in characteristics (columns (1) and (4)).

Columns (2) and (5) report the endowment terms for each group of variables: these are equivalent to the terms of the Gelbach decomposition, with the difference being that the female-male difference in characteristics is weighted by the male coefficient instead of the coefficient from the estimation on the pooled sample. The results indicate that the group of variables that contributes the most to the portion of the gap due to differences in endowments is the high school curriculum. Females are less represented in schools with higher returns to STEM/course-maths-intensity and more represented in schools with lower returns to STEM/course-maths-intensity, and this accounts for approximately half of the overall gender difference in outcomes. The endowment term related to the high school performance is positive and relevant in magnitude, indicating that if males performed as well as females in high school, the gender gap in the outcomes would be even larger.

Columns (3) and (6) report the coefficient terms for each group of variables. Most of the overall difference in coefficients is driven by different returns from the high school track and final grade: females have lower returns to high school tracks that are positively related to the choice of a STEM degree or of courses with higher maths content, and lower returns to a higher high school final grade.

To better understand which factors within each group of variables are driving the results of the Oaxaca decomposition, I report in table 2.5 the detailed decomposition for each variable within the most relevant groups – namely, high school track, family and social background for the STEM graduation rate and high school track, family characteristics and attitudes for the maths intensity measure. For both outcomes, most of the difference in endowments accounted for by the high school track variables is driven by a much lower rate at which females attend a scientific and a technical STEM high school. On the other hand, the difference in returns to the high school track is driven only by a lower probability of choosing a STEM major conditional on having attended a scientific high school.

The female-male difference in returns to family characteristics (columns (2) and (4)) is mostly accounted for by the variables measuring parents' occupation: from the full regression results performed separately for the samples of females and males reported in the appendix we observe that having a parent – in particular, the father – employed in a liberal profession has a negative correlation with the probability of choosing a STEM degree only for males. This result could be due to the fact that the son, not the daughter, in those families is more likely to follow the profession of the father (or of the mother) which are typically non-STEM occupations, such as doctors or lawyers.

For the STEM graduation outcome, I look at details for the variables measuring students' social background: the most significant term is the difference in the coefficients of the variable measuring the share of religious marriages. The full regression results indicate that this variable is negatively correlated with the probability of choosing a STEM degree for females but positively correlated for males. This result suggests that in societies that are less gender equal – as measured by at least one of the variables characterising attitudes towards women in a municipality – the gender gap in the major choices is even higher.

For the maths intensity of the course, I examine details regarding the role of attitudes in explaining the gender gap: the most relevant variables are the importance of career prospects – valued less by females – and culture – valued more by females. Moreover, even assuming that females and males give the same value to career prospects, I find that females have lower probability of choosing a course with higher maths intensity.

2.5.2 Sub-sample Analysis

In this section I investigate potential heterogeneity of the results across sub-samples defined according to the socio-economic status of the students' family.

The variable on socio-economic status is constructed based on the answers of students to questions regarding their parents' last occupation[10]. Through this step, three different social groups can be distinguished: low – parents in blue-collar jobs; medium – parents who are small business owners or low-level white-collar workers; high – parents who are directors or owners of businesses with at least 15 workers or who are self-employed in liberal professions.

Tables 2.6 and 2.7 present results from, respectively, the Gelbach and Oaxaca decompositions of the gender gap in STEM graduation rates for the three sub-samples. It emerges that the lower the socio-economic status is, the higher the raw gender gap, ranging from 16 percentage points for students belonging to the highest social class to 26 percentage points for students belonging to families where parents are blue-collar workers. This result is mainly driven by the fact that females' probability of graduating from STEM programmes increases with social status while the

[10]Following Schizzerotto (1994), the social class of the family refers to the highest between the two parents.

opposite is true for males – as shown in table 2.7 that reports the STEM graduation rates by gender. This evidence may be consistent with the hypothesis that in families where the parents are employed in liberal professions the male sons tend to follow the profession of the parents, which are typically non-STEM professions.

While the gender gap in major choices declines with socio-economic status, the role of the different groups of variables in explaining the gap does not appear to differ significantly across the three sub-samples. Table 2.6 shows that the high school track explains half of the gap in each sub-sample, and except for the high school performance, the other groups of variables always have negligible roles. The results from the Oaxaca decomposition, presented in table 2.7, are also fairly homogeneous across the different sub-samples: most of the unexplained portion of the gender gap is accounted for by lower returns of the high school track and performance for females.

The role of the high school experience as a main determinant of the different college choices of males and females is remarkably stable across social classes. This result is not completely unexpected, considering that the Italian high school system is characterised by a completely free access, such that a high level of segregation based on socio-economic status is not expected.

2.6 Conclusions

Despite the striking reversal of the gender gap in education in industrialised countries in the past 40 years, women pursue STEM degrees much less than their male peers do.

This paper assesses the relative importance of various explanations for the gender gap in STEM graduation rates for Italian college graduates. The major choices of students graduating from 2010 to 2015 are studied by exploiting a uniquely rich dataset obtained from the inter-university consortium AlmaLaurea. This dataset allows the measurement of students' high school experience, their attitudes and aspirations, and their family background. It is complemented with information on Italian municipalities from which I obtain measures of a student's sociocultural background characteristics, and with data on the local supply of degree programmes.

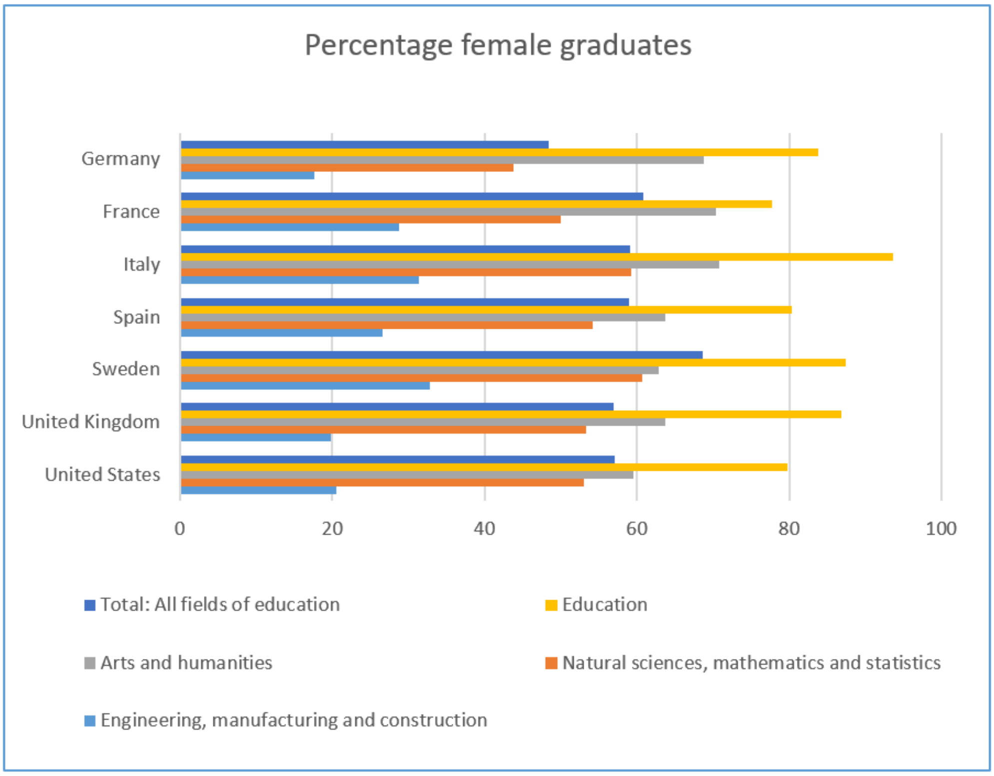
I evaluate the competing role of the different groups of variables and find that students' high school experience explains up to half of the gender gap in STEM graduation rates. Most of this is related to educational choices undertaken at an earlier stage, when young students choose between maths-intensive or humanities-oriented high school tracks. Young girls are less likely to choose tracks with a focus on maths and technical skills; this tends to refer, in particular, to the scientific 'Liceo' and the technical 'Istituto' with a focus on industrial construction and preparation for surveyors, which are the fields that ensure the highest returns to STEM enrolment in college. Even conditional on the high school track choice, a relevant role is played by a different influence of the family and social backgrounds on the decisions of females and males. Furthermore, my evidence demonstrates that females have attitudes suggesting lower competitiveness and higher altruism and social mindedness, which are negatively associated with the choice of a STEM degree, although these differences do not play a big role in explaining the gap in major choice.

By showing that high school track choices explain a large portion of the gender gap in STEM graduation, my results indicate that in Italy this issue has its roots in a gendered choice that has already taken place many years before. This finding suggests that the role of the influence of environmental factors – such as the family – in the different educational choices of females and males is even greater than can be estimated through this study.

These results have important policy implications. The findings indicate that effective interventions aimed at increasing girls' interests in science and technology should be implemented at an early stage, even in middle school, because the decision made by girls at 14 years of age will determine to a large extent their future education path and, consequently, their career and wage.

Figures and Tables

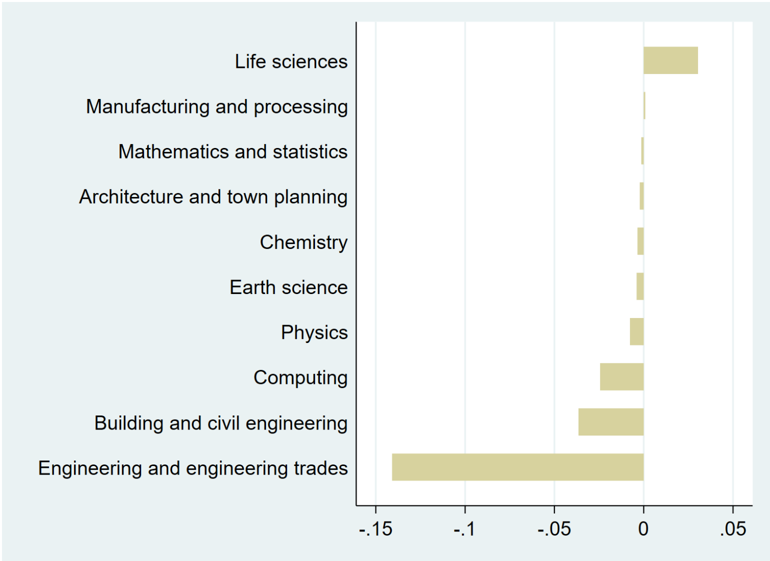
Figure 2.1: Gender differences in fields of study



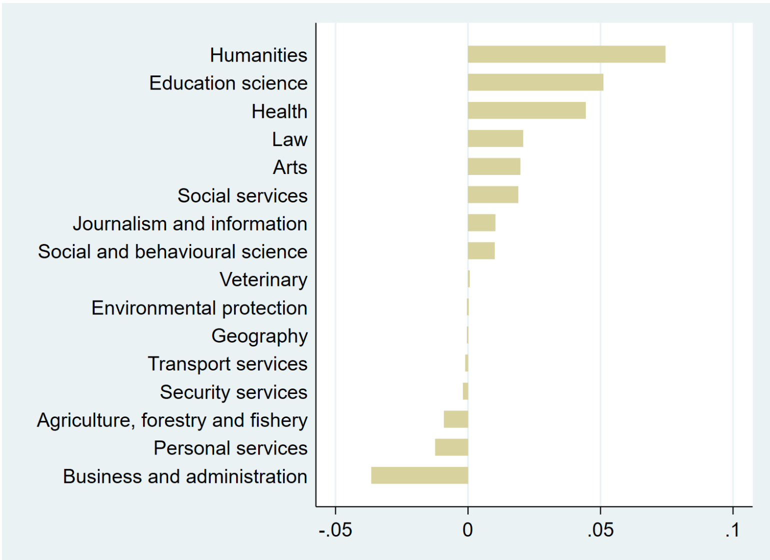
Source: OECD (2015b)

Figure 2.2: Enrolment gender gap in fields of study

(a) STEM fields



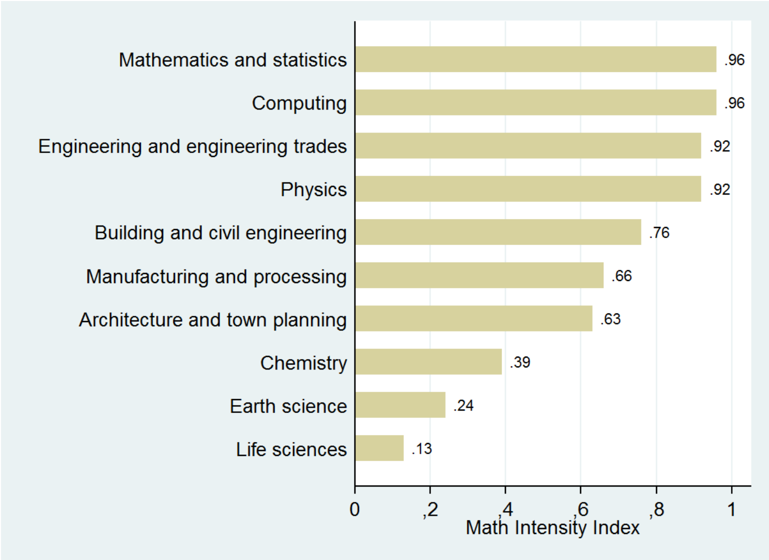
(b) non-STEM fields



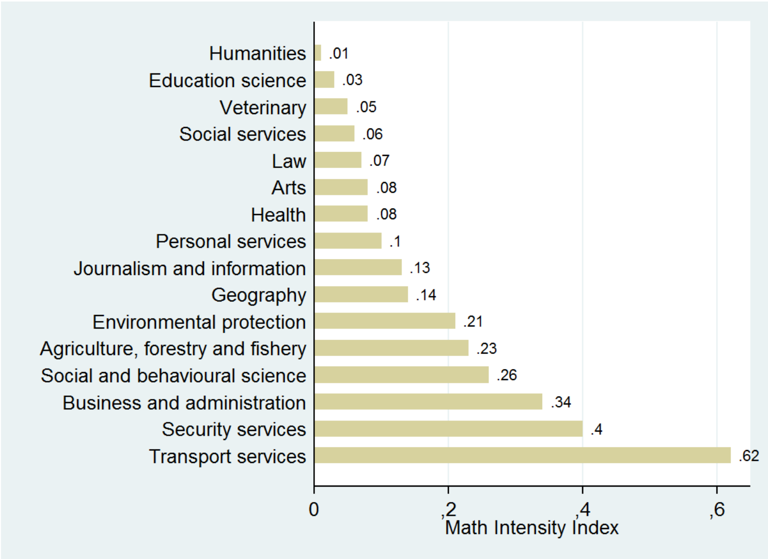
Notes: The figure plots the average female-male difference in enrolment probabilities for each group of university fields of study according to the FOET 1999 definition. Data are made available by the MIUR and are relative to the 2010/2011 academic year.

Figure 2.3: Maths intensity of fields of study

(a) STEM fields

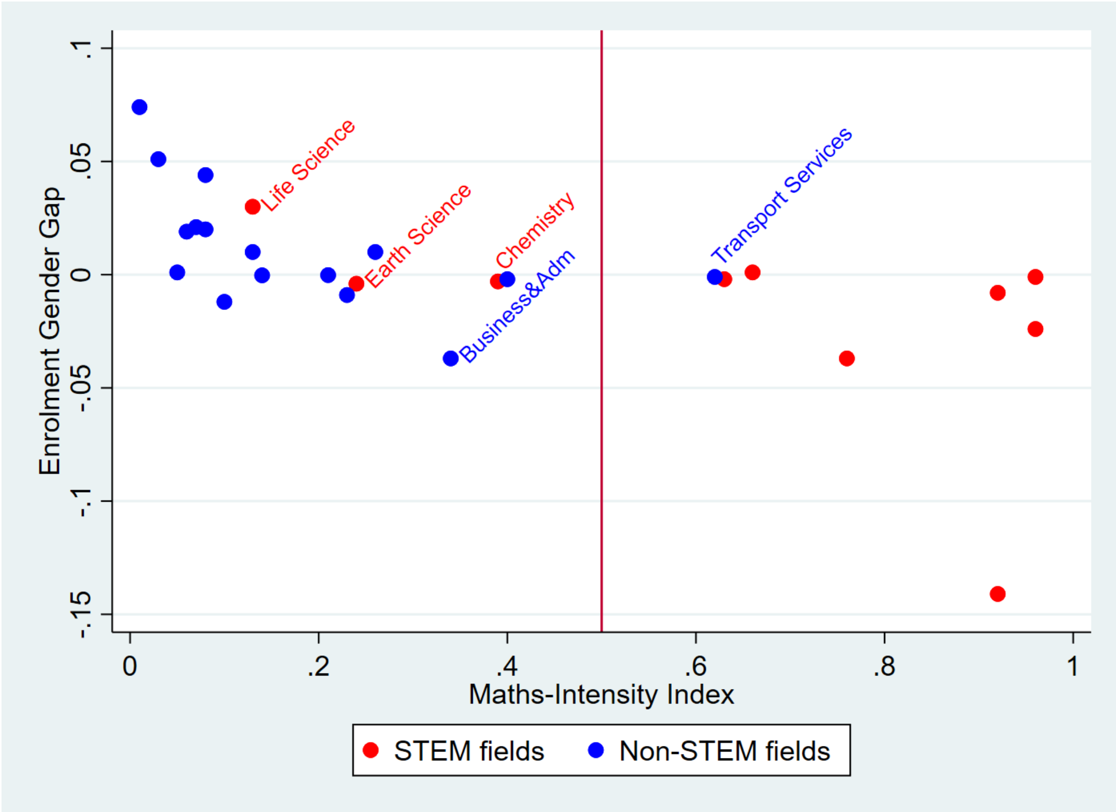


(b) non-STEM fields



Notes: The maths-intensity index is calculated as the percentage of college credits related to maths-intensive subjects out of the total credits for each field of study, averaged across all courses in a given field. Data are relative to the courses offered in the academic year 2010/2011.

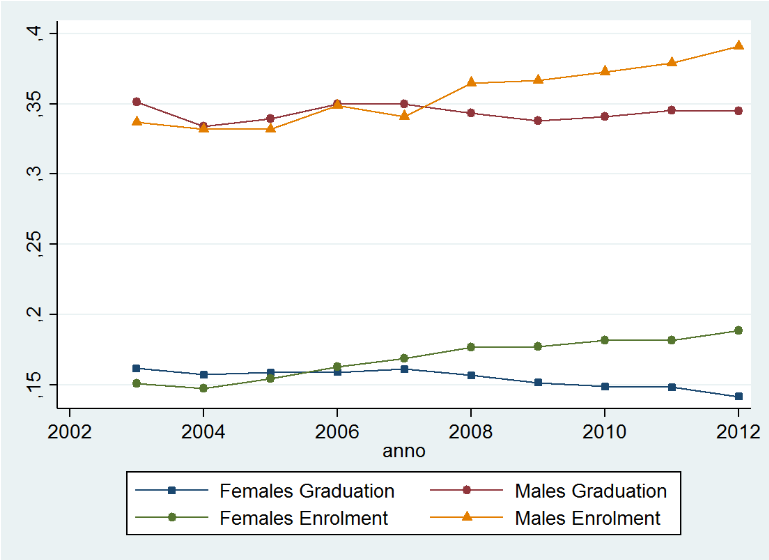
Figure 2.4: Enrolment gender gap and maths intensity by fields of education



Notes: Each observation is a field of study. The average maths intensity across all courses in a given field is represented on the x-axis, while the y-axis shows the female-male difference in the probability of enrolling in each field.

Figure 2.5: Enrolment and graduation rates in STEM fields

(a)



(b)

Notes: Enrolment rates (number of students enrolled in STEM fields as a percentage of the total number of students enrolled) are obtained from MIUR data for students enrolled in an undergraduate or single-cycle master's degree between 2003 and 2012 in universities taking part in the AlmaLaurea survey from 2010. Graduation rates (number of students graduated from STEM fields as a percentage of the total number of graduates) are obtained from AlmaLaurea data for students who graduated from an undergraduate or single-cycle master's degree programme between 2010 and 2015 and who enrolled between 2003 and 2012, from universities taking part in the AlmaLaurea survey from 2010.

Figure 2.6: Municipal variables

(a) Female Mayors

(b) Share of Female Councillors

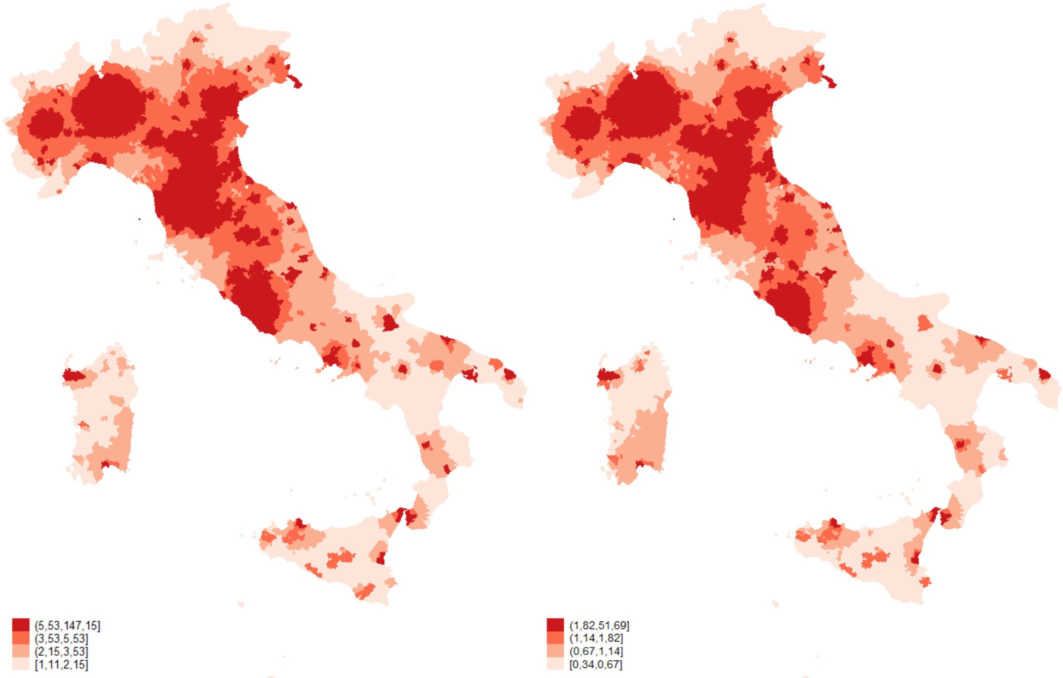
(c) Fertility Rate

(d) Share of Religious Marriages

Notes: All variables are measured in 2010. Panel (a) shows in red the municipalities governed by a female mayor, and panel (b) plots the share of female councillors in the local government at the municipal level. Both variables are obtained from data on local administrators from the Italian Ministry of the Interior. Panels (c) and (d) plot respectively the fertility rate – i.e., the ratio of the number of live births to the number of females aged 15-49 (times 1,000) – and the percentage of religious marriages, both obtained from the ISTAT *Atlante Statistico dei Comuni*.

Figure 2.7: Supply index

(a) All Courses (b) STEM courses



Notes: The two panels plot the index of supply in 2010, obtained for each municipality by summing the number of all/STEM-only courses offered in all other municipalities, weighted by the linear distance, the size of the university offering the course and the percentage of scholarships awarded by each university.

Table 2.1: FOET 1999 Classification

Broad fields	Fields of Education
1. Education	Teacher training and education science
2. Humanities and Arts	Arts
	Humanities
3. Social sciences, business and law	Social and behavioural science
	Journalism and information
	Business and administration
	Law
4. Science, Mathematics and Computing	**Life sciences**
	Physical sciences
	Mathematics and Statistics
	Computing
5. Engineering, Manufacturing and Construction	**Engineering and engineering trades**
	Manufacturing and processing
	Architecture and building
6. Agriculture	Agriculture, forestry and fishery
	Veterinary
7. Health and Welfare	Health
	Social services
8. Services	Personal services
	Transport services
	Environmental protection
	Security services

Notes: Source: Fields of Training Manual, European Centre for the Development of Vocational Training 1999

Table 2.2: Summary Statistics

Variables	Males		Females	
Observations	184,293		301,057	
	Mean	sd	Mean	sd
Stem	0.39	0.49	0.17	0.38
Maths intensity	0.41	0.35	0.20	0.25
High School:				
Humanities	0.18	0.38	0.47	0.50
Scientific & Technical	0.83	0.38	0.53	0.50
Final grade	80.7	12.4	83.7	12
Attitudes				
Enrolment motivation (professional)	0.12	0.33	0.08	0.28
Salary very important	0.57	0.50	0.57	0.50
Career prospects very important	0.66	0.47	0.61	0.49
Stability very important	0.65	0.48	0.75	0.43
Culture very important	0.38	0.49	0.46	0.50
Free time very important	0.26	0.44	0.26	0.44
Volunteering activities	0.21	0.41	0.25	0.43
Family Characteristics				
Father education:				
Less than HS	0.29	0.46	0.38	0.48
HS	0.46	0.50	0.44	0.50
College non STEM	0.17	0.37	0.13	0.34
College Science	0.02	0.14	0.02	0.12
College Engineering	0.06	0.23	0.04	0.20
Mother education:				
Less than HS	0.27	0.45	0.35	0.48
HS	0.51	0.50	0.48	0.50
College non STEM	0.18	0.38	0.14	0.35
College Science	0.03	0.18	0.02	0.15
College Engineering	0.01	0.09	0.01	0.08
Father last occupation:				
Blue collar (or never worked)	0.27	0.44	0.31	0.46
Self employed/small business owner	0.19	0.39	0.22	0.42
White collar	0.30	0.46	0.27	0.45
Liberal professions/entrepreneur	0.24	0.43	0.19	0.40
Mother last occupation:				
Housewife	0.23	0.42	0.26	0.44
Blue collar	0.28	0.45	0.29	0.45
Self employed/small business owner	0.10	0.30	0.11	0.31
White collar	0.32	0.47	0.29	0.45
Liberal professions/entrepreneur	0.07	0.26	0.06	0.24
Municipality Characteristics				
Fertility Rate	39.23	7.19	39.08	7.47
Religious marriages share	0.63	0.19	0.64	0.19
Female mayor	0.08	0.27	0.08	0.27
Share female councillors	0.14	0.10	0.14	0.10
Supply of STEM courses	7.8	16.0	6.9	15.0
Supply of university courses	24.5	49.7	21.7	46.6

Notes: Sample includes 3-year undergraduate or 5-year single-cycle students who enrolled between 2003 and 2011.

Table 2.3: Gelbach Coefficient Decomposition

Outcome:	STEM			Maths intensity		
	(1)	(2)	(3)	(4)	(5)	(6)
Estimated STEM gender gap	-0.219***			-0.218***		
	(0.00333)			(0.00228)		
HS curriculum:						
2 categories	-0.0455***			-0.0411***		
	(0.000386)			(0.000348)		
8 categories		-0.0982***			-0.0681***	
		(0.00188)			(0.00128)	
High school fixed effects			-0.103***			-0.0753***
			(0.00206)			(0.00136)
HS performance	0.0116***	0.0118***	0.0118***	0.00885***	0.00870***	0.00869***
	(0.000209)	(0.000212)	(0.000212)	(0.000160)	(0.000158)	(0.000158)
Attitudes	-0.00373***	-0.00441***	-0.00421***	-0.0114***	-0.0115***	-0.0110***
	(7.47e-05)	(8.24e-05)	(8.10e-05)	(0.000164)	(0.000165)	(0.000160)
Parents	-0.00414***	-0.00257***	-0.00231***	-0.00211***	-0.00238***	-0.00204***
	(0.000191)	(0.000175)	(0.000167)	(0.000119)	(0.000120)	(0.000113)
Municipal variables	3.30e-06	-5.87e-05***	-1.30e-05*	-6.24e-05***	-8.58e-05***	-3.95e-05***
	(9.43e-06)	(9.75e-06)	(7.47e-06)	(1.91e-05)	(2.03e-05)	(1.53e-05)
Supply	0.00131**	0.00134**	0.00152**	0.00101**	0.00101**	0.00109**
	(0.000578)	(0.000592)	(0.000674)	(0.000447)	(0.000446)	(0.000480)
Cohort fe	-8.83e-05***	0.000110***	0.000118***	0.000827***	0.00104***	0.00107***
	(1.40e-05)	(1.49e-05)	(1.47e-05)	(4.29e-05)	(5.35e-05)	(5.48e-05)
Municipality FE	-0.00281***	-0.00243***	-0.00209**	-0.00303***	-0.00281***	-0.00201***
	(0.000823)	(0.000800)	(0.000947)	(0.000623)	(0.000626)	(0.000736)
Full regression coefficient	-0.176***	-0.125***	-0.121***	-0.171***	-0.144***	-0.138***
	(0.00400)	(0.00239)	(0.00237)	(0.00253)	(0.00173)	(0.00170)
Observations	485,350	485,350	485,350	485,350	485,350	485,350
R squared	0.143	0.203	0.244	0.235	0.260	0.304

Notes: Decompositions of the gender gap in STEM graduation rate/maths intensity of university courses based on Gelbach (2016). The sample consists of college graduates who enrolled between 2003 and 2010 and graduated between 2010 and 2015. The dependent variable is a dummy equal to 1 if the individual graduated from a STEM field in columns (1)-(3) and the maths intensity of the course of study in columns (4)-(6). Each regression includes the survey year, year of graduation and municipality of residence fixed effects. The other variables are defined as follows. High school curriculum: 2 dummies for scientific/technical versus humanities in columns (1) and (4); 8 dummies for classics, education, languages, arts, technical non-STEM, technical STEM, science, and professional high school track in columns (2) and (5); more than 11,000 identifiers for secondary institution and track attended in columns (3) and (6). High school performance: 3 dummies for the intervals 60-85, 85-95, and 95-100. Attitudes: dummy=1 if the motivation to enrol in a course of study is professional versus cultural; dummies=1 if salary/career prospects/stability/culture/free time is very important versus slightly or not important in a future job; dummy=1 if engaged in volunteering activities. Parent characteristics: 5 dummies for father/mother's level of education (less than high school, high school, college non-STEM, college STEM science, and college STEM engineering); 4 dummies for father's last occupation (never worked or blue collar, small business man, white collar, liberal professions); and 5 dummies for mother's last occupation (housewife, blue collar, small business woman, white collar, liberal professions). Municipal variables: all variables measured in the municipality of residence in the year of university enrolment: dummy=1 if the mayor is female, share of female councillors, fertility rate, and share of religious marriages. Supply: indexes measuring the supply of STEM or overall university courses in the year of enrolment.

Table 2.4: Oaxaca Decomposition

Outcome:	STEM			Maths intensity		
	(1)	(2)	(3)	(4)	(5)	(6)
	Overall	Endowments	Coefficients	Overall	Endowments	Coefficients
Females	0.173***			0.195***		
	(0.00252)			(0.00189)		
Males	0.392***			0.413***		
	(0.00262)			(0.00210)		
Gender Gap	-0.219***			-0.218***		
	(0.00269)			(0.00188)		
Endowments	-0.0987***			-0.0853***		
	(0.00204)			(0.00146)		
Coefficients	-0.121***			-0.132***		
	(0.00234)			(0.00183)		
High School Track		-0.109***	-0.0477***		-0.0831***	-0.0405***
		(0.00208)	(0.00311)		(0.00148)	(0.00228)
High School performance		0.0177***	-0.0553***		0.0128***	-0.0389***
		(0.000442)	(0.00131)		(0.000322)	(0.000922)
Attitudes		-0.00594***	-0.000470		-0.0137***	0.00981***
		(0.000440)	(0.00257)		(0.000369)	(0.00185)
Family Characteristics		-0.00104***	0.0196***		-0.00131***	0.0107***
		(0.000359)	(0.00343)		(0.000298)	(0.00248)
Municipal Variables		4.84e-06	-0.0284*		4.87e-06	-0.00276
		(8.56e-06)	(0.0157)		(8.51e-06)	(0.0112)
Supply indexes		-3.72e-05	0.00545**		-3.80e-05	0.00509***
		(4.58e-05)	(0.00262)		(4.55e-05)	(0.00189)
Constant			-0.0137			-0.0758***
			(0.0171)			(0.0121)
Observations	485,350	485,350	485,350	485,350	485,350	485,350

Notes: Oaxaca decompositions of the gender gap in STEM graduation rate/maths intensity of university courses. The sample consists of college graduates who enrolled between 2003 and 2010 and graduated between 2010 and 2015. The dependent variable is a dummy equal to 1 if the individual graduated from a STEM field in columns (1)-(3) and the maths intensity of the course of study in columns (4)-(6). Each regression includes the survey year, year of graduation and municipality of residence fixed effects. The other variables are defined as in table 2.3.

Table 2.5: Detailed Oaxaca Decomposition

Outcome:		STEM				Maths intensity	
		(1)	(2)			(3)	(4)
VARIABLES		Endowments	Coefficients	VARIABLES		Endowments	Coefficients
High School Track	Overall	-0.109***	-0.0477***	High School Track	Overall	-0.0831***	-0.0405***
		(0.00208)	(0.00311)			(0.00148)	(0.00228)
	Education	-0.00555***	-0.00178*		Education	-0.00408***	-0.000618
		(0.000940)	(0.00102)			(0.000726)	(0.000755)
	Languages	-0.00433***	-0.00166**		Languages	-0.00303***	0.000279
		(0.000723)	(0.000844)			(0.000622)	(0.000693)
	Arts	0.00446***	-0.00311***		Arts	0.00233***	-0.00116***
		(0.000258)	(0.000330)			(0.000144)	(0.000200)
	Technical non STEM	-0.000941***	-0.00115*		Technical non STEM	0.00150***	-0.00121***
		(0.000123)	(0.000610)			(0.000167)	(0.000450)
	Technical STEM	-0.0651***	-0.00187***		Technical STEM	-0.0490***	-0.00213***
		(0.00276)	(0.000254)			(0.00207)	(0.000247)
	Science	-0.0379***	-0.0378***		Science	-0.0308***	-0.0357***
		(0.00160)	(0.00159)			(0.00128)	(0.00108)
	Professional	5.08e-06	-0.000328**		Professional	-1.11e-05	1.13e-05
		(7.61e-06)	(0.000161)			(8.14e-06)	(0.000108)
Family Characteristics	Overall	-0.00104***	0.0196***	Family Characteristics	Overall	-0.00131***	0.0107***
		(0.000359)	(0.00343)			(0.000298)	(0.00248)
	Parents education	-0.00167***	0.00613**		Parents education	-0.00153***	0.00312
		(0.000395)	(0.00279)			(0.000320)	(0.00206)
	Parents last occupation	0.000629***	0.0134***		Parents last occupation	0.000219*	0.00762***
		(0.000167)	(0.00320)			(0.000127)	(0.00223)
Municipal Variables	Overall	4.84e-06	-0.0284*	Attitudes	Overall	-0.0137***	0.00981***
		(8.56e-06)	(0.0157)			(0.000369)	(0.00185)
	Female mayor	1.08e-06	0.000112		Enrolment motivation (professional)	-0.00316***	-0.000198
		(5.23e-06)	(0.000753)			(0.000122)	(0.000245)
	Share female councillors	4.72e-07	0.00235		Salary very important	-9.87e-06	1.99e-05
		(2.36e-06)	(0.00325)			(8.62e-06)	(0.00137)
	Fertility rate	1.35e-06	-0.00842		Career prospects very important	-0.00317***	-0.0127***
		(4.17e-06)	(0.0115)			(0.000175)	(0.00134)
	Share of religious marriages	1.94e-06	-0.0225**		Stability very important	-0.000852***	0.00104
		(4.97e-06)	(0.00974)			(0.000175)	(0.00179)
					Culture very important	-0.00512***	0.0136***
						(0.000200)	(0.00109)
					Free time very important	-7.18e-05*	0.00482***
						(4.28e-05)	(0.000571)
					Volunteering activities	-0.00129***	0.00323***
						(9.05e-05)	(0.000557)

Notes: Details of the Oaxaca decomposition results presented in table 2.4. The table presents in columns (1) and (2) the endowment and coefficient terms of the gender gap in STEM graduation for the different variables within the groups: high school track (8 categories), family characteristics (parents' education and parents' last occupation), and municipal variables (female mayor, share of female councillors, fertility rate, share of religious marriages). In columns (3) and (4) the table presents the endowment and coefficient terms of the gender gap in maths intensity of the course of study for the different variables within the groups: high school track and family characteristics as in the other columns, and attitudes (enrolment motivation, importance of salary/career/stability/culture/free time for future jobs, involvement in volunteering activities.)

Table 2.6: Gelbach Decomposition by Socio-economic Status

Socio-economic status:	High			Medium			Low		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Estimated STEM gender gap	-0.158***			-0.229***			-0.264***		
	(0.00416)			(0.00312)			(0.00337)		
HS curriculum									
	-0.0424***			-0.0485***			-0.0469***		
	(0.000539)			(0.000451)			(0.000463)		
		-0.0655***			-0.102***			-0.122***	
		(0.00176)			(0.00156)			(0.00158)	
			-0.0661***			-0.107***			-0.130***
			(0.00215)			(0.00201)			(0.00213)
HS performance	0.0125***	0.0124***	0.0123***	0.0123***	0.0125***	0.0125***	0.00771***	0.00818***	0.00826***
	(0.000316)	(0.000312)	(0.000309)	(0.000309)	(0.000314)	(0.000315)	(0.000403)	(0.000427)	(0.000432)
Attitudes	-0.00333***	-0.00370***	-0.00365***	-0.00375***	-0.00451***	-0.00433***	-0.00457***	-0.00533***	-0.00520***
	(0.000137)	(0.000145)	(0.000148)	(9.76e-05)	(0.000108)	(0.000106)	(0.000169)	(0.000187)	(0.000185)
Parents	-0.00265***	-0.00225***	-0.00209***	-0.00362***	-0.00238***	-0.00219***	-0.00244***	-0.00154***	-0.00136***
	(0.000487)	(0.000467)	(0.000454)	(0.000172)	(0.000145)	(0.000137)	(0.000114)	(9.17e-05)	(8.49e-05)
Municipal variables	-0.000105**	-0.000105***	-6.45e-05	9.13e-05***	3.33e-05**	8.54e-05***	-5.43e-06	-5.03e-05***	-2.37e-05**
	(4.28e-05)	(3.98e-05)	(4.00e-05)	(1.71e-05)	(1.42e-05)	(2.00e-05)	(2.57e-05)	(1.71e-05)	(1.12e-05)
Supply	0.000589**	0.000579**	0.000621***	0.000753**	0.000794**	0.00102**	0.000715*	0.000822*	0.000821*
	(0.000244)	(0.000244)	(0.000236)	(0.000347)	(0.000367)	(0.000470)	(0.000378)	(0.000435)	(0.000435)
Cohort fe	0.00226***	0.00253***	0.00251***	-0.000176***	9.70e-06	2.85e-05**	-0.00188***	-0.00181***	-0.00167***
	(0.000220)	(0.000243)	(0.000241)	(2.14e-05)	(1.57e-05)	(1.37e-05)	(0.000159)	(0.000152)	(0.000143)
Municipality FE	-0.00161**	-0.00136*	-0.00166	-0.00255***	-0.00220***	-0.00145*	-0.00271***	-0.00210***	-0.000272
	(0.000751)	(0.000729)	(0.00118)	(0.000642)	(0.000609)	(0.000862)	(0.000793)	(0.000778)	(0.000983)
Full regression coefficient	-0.123***	-0.100***	-0.0993***	-0.184***	-0.132***	-0.128***	-0.214***	-0.140***	-0.134***
	(0.00515)	(0.00402)	(0.00406)	(0.00378)	(0.00256)	(0.00257)	(0.00367)	(0.00311)	(0.00332)
Observations	111,210	111,210	111,210	250,944	250,944	250,944	113,606	113,606	113,606
R squared	0.161	0.196	0.253	0.153	0.215	0.265	0.192	0.264	0.333

Notes: Decompositions of the gender gap in STEM graduation based on Gelbach (2016) for three sub-samples defined according to the socio-economic status of the students' family (high/medium/low). For each sub-sample, three models with different definitions of high school tracks are estimated: 2 dummies for scientific/technical versus humanities in columns (1),(4) and (7); 8 dummies for classics, education, languages, arts, technical non-STEM, technical STEM, science, and professional high school track in columns (2),(5) and (8); more than 11,000 identifiers for the secondary institution and track attended in columns (3),(6) and (9). The other variables are defined as in table 2.3.

Table 2.7: Sub-sample Analysis: Oaxaca Decomposition

Socio-economic status:	High			Medium			Low		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Overall	Explained	Unexplained	Overall	Explained	Unexplained	Overall	Explained	Unexplained
Females	0.200***			0.176***			0.146***		
	(0.00410)			(0.00236)			(0.00198)		
Males	0.358***			0.406***			0.410***		
	(0.00368)			(0.00252)			(0.00331)		
Gender Gap	-0.158***			-0.229***			-0.264***		
	(0.00361)			(0.00279)			(0.00338)		
Endowments	-0.0590***			-0.0997***			-0.128***		
	(0.00224)			(0.00213)			(0.00337)		
Coefficients	-0.0986***			-0.130***			-0.136***		
	(0.00398)			(0.00271)			(0.00375)		
High School Track		-0.0705***	-0.0335***		-0.112***	-0.0587***		-0.131***	-0.0900***
		(0.00223)	(0.00400)		(0.00205)	(0.00469)		(0.00323)	(0.00820)
High School performance		0.0183***	-0.0462***		0.0187***	-0.0584***		0.0126***	-0.0611***
		(0.000801)	(0.00268)		(0.000632)	(0.00185)		(0.000759)	(0.00266)
Attitudes		-0.00461***	0.00248		-0.00549***	-0.00494		-0.00780***	0.00498
		(0.000788)	(0.00486)		(0.000544)	(0.00368)		(0.000780)	(0.00562)
Family Characteristics		-0.00203***	0.00245		-0.00133***	0.0126***		-0.00139***	-0.000640
		(0.000640)	(0.0142)		(0.000329)	(0.00322)		(0.000422)	(0.00280)
Municipal Variables		-6.90e-06	-0.0542		1.27e-07	-0.0470**		-9.15e-06	-0.0267
		(2.11e-05)	(0.0404)		(2.08e-05)	(0.0224)		(4.30e-05)	(0.0338)
Supply indexes		-6.84e-05	0.0135*		-2.00e-05	0.000599		1.34e-05	0.00944***
		(5.98e-05)	(0.00695)		(4.61e-05)	(0.00415)		(4.76e-05)	(0.00307)
Constant			0.0169			0.0260			0.0280
			(0.0437)			(0.0239)			(0.0354)
Observations	111,210	111,210	111,210	250,944	250,944	250,944	113,606	113,606	113,606

Notes: Oaxaca decompositions of the gender gap in STEM graduation for three sub-samples defined according to the socio-economic status of the students' family (high/medium/low).

Appendices

Table A1: Full Regressions: STEM Graduation Rate

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
female	-0.176***			-0.125***			-0.121***		
	(0.00400)			(0.00239)			(0.00237)		
High School track (humanities excluded):									
Scientific/Technical	0.156***	0.236***	0.127***						
	(0.00236)	(0.00329)	(0.00229)						
Education				-0.0605***	-0.0965***	-0.0682***			
				(0.00324)	(0.00789)	(0.00311)			
Languages				-0.0557***	-0.0660***	-0.0614***			
				(0.00308)	(0.00680)	(0.00298)			
Arts				0.249***	0.364***	0.218***			
				(0.00802)	(0.0141)	(0.00838)			
Technical non STEM				-0.0705***	-0.0526***	-0.0650***			
				(0.00264)	(0.00476)	(0.00297)			
Technical STEM				0.396***	0.434***	0.371***			
				(0.00424)	(0.00574)	(0.00878)			
Science				0.179***	0.246***	0.148***			
				(0.00226)	(0.00431)	(0.00247)			
Professional				-0.0274***	-0.00449	-0.0273***			
				(0.00479)	(0.00918)	(0.00519)			
School dummies							YES	YES	YES
High school final grade:									
85-95	0.0916***	0.156***	0.0506***	0.0930***	0.152***	0.0553***	0.0921***	0.152***	0.0557***
	(0.00188)	(0.00321)	(0.00197)	(0.00190)	(0.00318)	(0.00195)	(0.00196)	(0.00329)	(0.00201)
95-100	0.140***	0.231***	0.0863***	0.143***	0.228***	0.0923***	0.143***	0.231***	0.0940***
	(0.00247)	(0.00391)	(0.00237)	(0.00244)	(0.00377)	(0.00234)	(0.00265)	(0.00407)	(0.00250)
Attitudes									
Enrolment motivation (professional)	0.0190***	0.0388***	-0.00121	0.0204***	0.0373***	0.00293	0.0196***	0.0355***	0.00262
	(0.00234)	(0.00347)	(0.00264)	(0.00222)	(0.00336)	(0.00255)	(0.00224)	(0.00351)	(0.00254)
Salary very important	0.00394***	-0.00133	0.00734***	0.00507***	0.000155	0.00740***	0.00445***	-0.00118	0.00722***
	(0.00143)	(0.00275)	(0.00200)	(0.00137)	(0.00265)	(0.00192)	(0.00144)	(0.00267)	(0.00205)
Career prospects very important	0.0127***	0.0194***	0.00454**	0.0147***	0.0212***	0.00830***	0.0148***	0.0222***	0.00813***
	(0.00160)	(0.00302)	(0.00177)	(0.00155)	(0.00285)	(0.00174)	(0.00146)	(0.00282)	(0.00173)
Stability very important	0.00554***	0.0161***	-0.00308	0.00623***	0.0141***	0.000254	0.00823***	0.0161***	0.00221
	(0.00156)	(0.00250)	(0.00213)	(0.00149)	(0.00236)	(0.00209)	(0.00141)	(0.00240)	(0.00194)
Culture very important	-0.0205***	-0.0353***	-0.0107***	-0.0277***	-0.0384***	-0.0203***	-0.0282***	-0.0403***	-0.0205***
	(0.00140)	(0.00289)	(0.00141)	(0.00138)	(0.00281)	(0.00138)	(0.00135)	(0.00286)	(0.00139)
Free time very important	-0.0237***	-0.0358***	-0.0159***	-0.0214***	-0.0345***	-0.0130***	-0.0195***	-0.0309***	-0.0120***
	(0.00153)	(0.00268)	(0.00167)	(0.00151)	(0.00256)	(0.00167)	(0.00156)	(0.00274)	(0.00175)
Volunteering activities	-0.0300***	-0.0400***	-0.0255***	-0.0304***	-0.0370***	-0.0266***	-0.0301***	-0.0369***	-0.0263***
	(0.00141)	(0.00277)	(0.00168)	(0.00136)	(0.00264)	(0.00162)	(0.00133)	(0.00263)	(0.00162)

cont'd

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Father education (less than HS excluded):									
High school	0.0154***	0.00794**	0.0192***	0.00885***	0.00897***	0.00929***	0.00779***	0.00805**	0.00767***
	(0.00158)	(0.00314)	(0.00165)	(0.00147)	(0.00296)	(0.00156)	(0.00152)	(0.00314)	(0.00161)
College non STEM	-0.0264***	-0.0571***	0.00233	-0.0319***	-0.0509***	-0.0138***	-0.0311***	-0.0499***	-0.0139***
	(0.00343)	(0.00524)	(0.00332)	(0.00321)	(0.00502)	(0.00321)	(0.00292)	(0.00491)	(0.00305)
College STEM Science	0.0903***	0.0752***	0.102***	0.0779***	0.0740***	0.0806***	0.0740***	0.0681***	0.0781***
	(0.00537)	(0.00819)	(0.00651)	(0.00531)	(0.00815)	(0.00646)	(0.00508)	(0.00780)	(0.00677)
College STEM Engineering	0.152***	0.167***	0.139***	0.135***	0.156***	0.115***	0.129***	0.150***	0.110***
	(0.00442)	(0.00627)	(0.00545)	(0.00471)	(0.00671)	(0.00547)	(0.00455)	(0.00726)	(0.00525)
Mother education (less than HS excluded):									
High school	0.00876***	-0.00488	0.0151***	0.00362**	0.000380	0.00553***	0.00332**	0.00129	0.00439**
	(0.00151)	(0.00310)	(0.00168)	(0.00146)	(0.00287)	(0.00177)	(0.00148)	(0.00299)	(0.00180)
College non STEM	0.0128***	0.000646	0.0237***	0.00571**	0.00587	0.00769***	0.00465*	0.00482	0.00577**
	(0.00282)	(0.00535)	(0.00267)	(0.00262)	(0.00500)	(0.00262)	(0.00245)	(0.00487)	(0.00267)
College STEM Science	0.0906***	0.0679***	0.107***	0.0788***	0.0709***	0.0844***	0.0739***	0.0671***	0.0779***
	(0.00545)	(0.00755)	(0.00681)	(0.00534)	(0.00731)	(0.00679)	(0.00475)	(0.00693)	(0.00637)
College STEM Engineering	0.124***	0.116***	0.133***	0.112***	0.115***	0.112***	0.105***	0.107***	0.103***
	(0.0104)	(0.0146)	(0.0117)	(0.0104)	(0.0150)	(0.0116)	(0.00901)	(0.0140)	(0.0110)
Father last occupation (blue collar or never worked excluded):									
Self-employed/small businessman	0.00398**	0.000862	0.00653***	0.00424**	0.00466	0.00448**	0.00340**	0.00333	0.00403**
	(0.00179)	(0.00348)	(0.00188)	(0.00171)	(0.00330)	(0.00183)	(0.00167)	(0.00325)	(0.00188)
White collar	0.0118***	0.00863***	0.0128***	0.00890***	0.00953***	0.00797***	0.00792***	0.00869***	0.00757***
	(0.00176)	(0.00300)	(0.00217)	(0.00174)	(0.00287)	(0.00214)	(0.00175)	(0.00303)	(0.00213)
Liberal professions/white collar director/entrepreneur	0.00168	-0.0158***	0.0153***	-0.00133	-0.0131***	0.00816***	-0.00166	-0.0141***	0.00764***
	(0.00214)	(0.00375)	(0.00278)	(0.00212)	(0.00357)	(0.00274)	(0.00216)	(0.00360)	(0.00290)
Mother last occupation (housewife excluded):									
Blue collar	-0.00285	-0.00662*	-0.000423	-0.00507***	-0.00920***	-0.00292	-0.00659***	-0.0118***	-0.00420**
	(0.00179)	(0.00358)	(0.00201)	(0.00173)	(0.00343)	(0.00196)	(0.00168)	(0.00348)	(0.00197)
Self-employed/small businessman	0.000380	-0.0160***	0.0108***	-0.000774	-0.0128***	0.00656**	-0.000416	-0.0105**	0.00556**
	(0.00247)	(0.00449)	(0.00285)	(0.00234)	(0.00420)	(0.00272)	(0.00235)	(0.00428)	(0.00275)
White collar	0.00573***	-0.00811**	0.0159***	0.00198	-0.00949***	0.0100***	0.000626	-0.0106***	0.00866***
	(0.00192)	(0.00384)	(0.00242)	(0.00185)	(0.00364)	(0.00238)	(0.00183)	(0.00362)	(0.00240)
Liberal professions/white collar director/entrepreneur	-0.0109***	-0.0222***	-0.000139	-0.0134***	-0.0221***	-0.00605*	-0.0133***	-0.0201***	-0.00748**
	(0.00286)	(0.00511)	(0.00351)	(0.00281)	(0.00487)	(0.00355)	(0.00268)	(0.00520)	(0.00337)
Municipal Variables									
Female mayor	0.00519	0.00516	0.00456	0.00317	0.00361	0.00284	0.00367	0.00218	0.00496
	(0.00546)	(0.0105)	(0.00439)	(0.00527)	(0.00949)	(0.00434)	(0.00515)	(0.00887)	(0.00455)
Share female councillors	0.00325	-0.00894	0.0140	0.00514	0.00182	0.0121	0.00531	0.00540	0.00940
	(0.0170)	(0.0250)	(0.0179)	(0.0165)	(0.0235)	(0.0175)	(0.0169)	(0.0246)	(0.0175)
Fertility rate	0.000117	0.000371	-5.75e-06	0.000170	0.000421	8.24e-05	0.000128	0.000489*	3.49e-05
	(0.000150)	(0.000288)	(0.000166)	(0.000144)	(0.000273)	(0.000160)	(0.000144)	(0.000284)	(0.000163)
Share of religious marriages	0.000826	0.0258*	-0.0108	-0.00362	0.0189	-0.0142*	-5.30e-05	0.0264*	-0.0132
	(0.00766)	(0.0147)	(0.00851)	(0.00742)	(0.0141)	(0.00825)	(0.00738)	(0.0144)	(0.00832)
Supply of STEM courses	-0.00379	-0.00589***	-0.00242	-0.00410	-0.00566***	-0.00290	-0.00373	-0.00601**	-0.00279
	(0.00252)	(0.00225)	(0.00348)	(0.00253)	(0.00212)	(0.00337)	(0.00286)	(0.00240)	(0.00338)
Supply of university courses	0.000776	0.00127*	0.000503	0.000865	0.00123*	0.000608	0.000677	0.00132*	0.000467
	(0.000744)	(0.000726)	(0.00109)	(0.000748)	(0.000668)	(0.00106)	(0.000820)	(0.000692)	(0.00106)
Observations	485,350	183,588	300,787	485,350	183,588	300,787	485,350	181,294	299,321
R-squared	0.143	0.142	0.086	0.203	0.211	0.135	0.244	0.270	0.181

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Notes: The sample consists of college graduates who were 18 between 2003 and 2010 and who graduated between 2010 and 2015. The dependent variable is a binary variable equal 1 if the individual graduated from a STEM field. Each regression includes the survey year, year of graduation and municipality of residence fixed effects.

Table A2: Full Regressions: Maths Intensity of University Courses

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
female	-0.171***			-0.144***			-0.138***		
	(0.00253)			(0.00173)			(0.00170)		
High School track (humanities excluded):									
Scientific/Technical	0.141***	0.207***	0.117***						
	(0.00150)	(0.00273)	(0.00141)						
Education				-0.0363***	-0.0710***	-0.0449***			
				(0.00239)	(0.00593)	(0.00226)			
Languages				-0.0257***	-0.0457***	-0.0308***			
				(0.00264)	(0.00599)	(0.00250)			
Arts				0.146***	0.191***	0.133***			
				(0.00475)	(0.00855)	(0.00509)			
Technical non STEM				0.0776***	0.0945***	0.0802***			
				(0.00262)	(0.00427)	(0.00286)			
Technical STEM				0.286***	0.329***	0.242***			
				(0.00308)	(0.00422)	(0.00606)			
Science				0.137***	0.200***	0.107***			
				(0.00164)	(0.00323)	(0.00160)			
Professional				0.00632*	0.0168**	0.0130***			
				(0.00358)	(0.00692)	(0.00367)			
School dummies									
High school final grade:									
85-95	0.0676***	0.111***	0.0402***	0.0662***	0.107***	0.0400***	0.0656***	0.107***	0.0403***
	(0.00127)	(0.00231)	(0.00128)	(0.00131)	(0.00229)	(0.00132)	(0.00135)	(0.00245)	(0.00133)
95-100	0.109***	0.168***	0.0736***	0.107***	0.164***	0.0730***	0.107***	0.166***	0.0738***
	(0.00169)	(0.00291)	(0.00160)	(0.00172)	(0.00287)	(0.00164)	(0.00190)	(0.00325)	(0.00171)
Attitudes									
Enrolment motivation (professional)	0.0833***	0.0834***	0.0808***	0.0841***	0.0833***	0.0818***	0.0822***	0.0804***	0.0815***
	(0.00190)	(0.00242)	(0.00237)	(0.00187)	(0.00237)	(0.00234)	(0.00184)	(0.00235)	(0.00234)
Salary very important	0.00396***	0.00203	0.00473***	0.00465***	0.00316*	0.00459***	0.00462***	0.00209	0.00509***
	(0.00114)	(0.00183)	(0.00153)	(0.00112)	(0.00177)	(0.00152)	(0.00116)	(0.00179)	(0.00160)
Career prospects very important	0.0452***	0.0546***	0.0362***	0.0444***	0.0538***	0.0362***	0.0428***	0.0529***	0.0343***
	(0.00118)	(0.00208)	(0.00131)	(0.00114)	(0.00199)	(0.00128)	(0.00104)	(0.00198)	(0.00125)
Stability very important	-0.00806***	-0.00630***	-0.00978***	-0.00840***	-0.00787***	-0.00861***	-0.00644***	-0.00548***	-0.00720***
	(0.00102)	(0.00177)	(0.00138)	(0.000976)	(0.00173)	(0.00133)	(0.000928)	(0.00168)	(0.00126)
Culture very important	-0.0456***	-0.0638***	-0.0336***	-0.0474***	-0.0640***	-0.0362***	-0.0462***	-0.0630***	-0.0348***
	(0.000988)	(0.00220)	(0.000922)	(0.000948)	(0.00213)	(0.000927)	(0.000932)	(0.00213)	(0.000963)
Free time very important	-0.0214***	-0.0311***	-0.0145***	-0.0212***	-0.0313***	-0.0139***	-0.0187***	-0.0276***	-0.0122***
	(0.00103)	(0.00189)	(0.00116)	(0.00102)	(0.00186)	(0.00118)	(0.00104)	(0.00194)	(0.00121)
Volunteering activities	-0.0258***	-0.0343***	-0.0216***	-0.0253***	-0.0326***	-0.0210***	-0.0247***	-0.0318***	-0.0205***
	(0.00104)	(0.00212)	(0.00103)	(0.00102)	(0.00207)	(0.00101)	(0.00103)	(0.00209)	(0.00104)

cont'd

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Father education (less than HS excluded):									
High school	0.0111***	0.00884***	0.0118***	0.0108***	0.0120***	0.0100***	0.00924***	0.0106***	0.00800***
	(0.00113)	(0.00224)	(0.00115)	(0.00108)	(0.00217)	(0.00112)	(0.00111)	(0.00231)	(0.00115)
College non STEM	-0.0206***	-0.0399***	-0.00133	-0.0174***	-0.0310***	-0.00412*	-0.0184***	-0.0329***	-0.00553***
	(0.00296)	(0.00463)	(0.00232)	(0.00259)	(0.00434)	(0.00211)	(0.00231)	(0.00426)	(0.00200)
College STEM Science	0.0413***	0.0389***	0.0440***	0.0424***	0.0445***	0.0403***	0.0387***	0.0393***	0.0378***
	(0.00450)	(0.00695)	(0.00469)	(0.00432)	(0.00681)	(0.00456)	(0.00379)	(0.00644)	(0.00438)
College STEM Engineering	0.0985***	0.108***	0.0899***	0.0972***	0.109***	0.0857***	0.0916***	0.102***	0.0808***
	(0.00331)	(0.00482)	(0.00368)	(0.00332)	(0.00506)	(0.00354)	(0.00307)	(0.00514)	(0.00343)
Mother education (less than HS excluded):									
High school	0.00228**	-0.00438**	0.00529***	0.00340***	0.00168	0.00447***	0.00312***	0.00283	0.00375***
	(0.000989)	(0.00207)	(0.00105)	(0.000963)	(0.00199)	(0.00107)	(0.000974)	(0.00206)	(0.00108)
College non STEM	0.00425**	-0.000509	0.00962***	0.00702***	0.00802**	0.00780***	0.00578***	0.00647*	0.00655***
	(0.00209)	(0.00377)	(0.00200)	(0.00182)	(0.00348)	(0.00189)	(0.00165)	(0.00336)	(0.00185)
College STEM Science	0.0524***	0.0432***	0.0589***	0.0552***	0.0520***	0.0568***	0.0505***	0.0471***	0.0529***
	(0.00394)	(0.00523)	(0.00469)	(0.00370)	(0.00508)	(0.00446)	(0.00319)	(0.00470)	(0.00416)
College STEM Engineering	0.0777***	0.0792***	0.0788***	0.0778***	0.0848***	0.0744***	0.0723***	0.0775***	0.0684***
	(0.00637)	(0.00869)	(0.00749)	(0.00604)	(0.00863)	(0.00717)	(0.00514)	(0.00833)	(0.00655)
Father last occupation (blue collar or never worked excluded):									
Self-employed/small businessman	0.0127***	0.0111***	0.0143***	0.0136***	0.0139***	0.0138***	0.0123***	0.0140***	0.0120***
	(0.00127)	(0.00253)	(0.00129)	(0.00126)	(0.00245)	(0.00131)	(0.00126)	(0.00249)	(0.00133)
White collar	0.0102***	0.0100***	0.00939***	0.0104***	0.0120***	0.00893***	0.00932***	0.0114***	0.00791***
	(0.00115)	(0.00213)	(0.00131)	(0.00116)	(0.00208)	(0.00134)	(0.00118)	(0.00229)	(0.00132)
Liberal professions/white collar director/entrepreneur	0.00754***	-0.00130	0.0150***	0.00822***	0.00220	0.0135***	0.00727***	0.00176	0.0120***
	(0.00142)	(0.00272)	(0.00180)	(0.00143)	(0.00259)	(0.00184)	(0.00149)	(0.00260)	(0.00198)
Mother last occupation (housewife excluded):									
Blue collar	-0.00110	-0.00192	-0.000787	-0.00142	-0.00233	-0.00129	-0.00199*	-0.00390*	-0.00124
	(0.00121)	(0.00237)	(0.00128)	(0.00119)	(0.00230)	(0.00126)	(0.00114)	(0.00229)	(0.00127)
Self-employed/small businessman	0.00275*	-0.00890***	0.0100***	0.00322**	-0.00594**	0.00884***	0.00397**	-0.00350	0.00846***
	(0.00164)	(0.00308)	(0.00185)	(0.00158)	(0.00296)	(0.00179)	(0.00156)	(0.00305)	(0.00181)
White collar	0.000641	-0.00747***	0.00668***	0.000706	-0.00634**	0.00581***	0.000520	-0.00591**	0.00521***
	(0.00130)	(0.00263)	(0.00140)	(0.00129)	(0.00259)	(0.00136)	(0.00126)	(0.00263)	(0.00138)
Liberal professions/white collar director/entrepreneur	-0.00758***	-0.0172***	0.00119	-0.00731***	-0.0155***	-0.000504	-0.00778***	-0.0141***	-0.00252
	(0.00213)	(0.00373)	(0.00245)	(0.00208)	(0.00369)	(0.00237)	(0.00204)	(0.00418)	(0.00226)
Municipal Variables									
Female mayor	0.00888	0.0104	0.00703	0.00843	0.00964	0.00722	0.00692	0.00717	0.00704
	(0.00589)	(0.00903)	(0.00469)	(0.00574)	(0.00843)	(0.00482)	(0.00533)	(0.00789)	(0.00460)
Share female councillors	0.0121	0.00555	0.0199*	0.0127	0.0103	0.0186*	0.0120	0.0141	0.0165
	(0.0112)	(0.0180)	(0.0113)	(0.0111)	(0.0175)	(0.0113)	(0.0113)	(0.0180)	(0.0114)
Fertility rate	5.36e-05	8.37e-05	4.64e-05	7.57e-05	0.000101	8.70e-05	1.81e-05	0.000123	3.46e-05
	(0.000101)	(0.000198)	(0.000111)	(9.91e-05)	(0.000193)	(0.000109)	(9.81e-05)	(0.000199)	(0.000109)
Share of religious marriages	-0.00686	0.00319	-0.0111**	-0.00860*	0.000272	-0.0123**	-0.00470	0.00323	-0.00942*
	(0.00517)	(0.0102)	(0.00549)	(0.00507)	(0.00992)	(0.00542)	(0.00502)	(0.0101)	(0.00545)
Supply of STEM courses	-0.00422***	-0.00644***	-0.00259	-0.00408**	-0.00621***	-0.00235	-0.00360*	-0.00596***	-0.00211
	(0.00162)	(0.00142)	(0.00227)	(0.00169)	(0.00136)	(0.00230)	(0.00195)	(0.00159)	(0.00236)
Supply of university courses	0.00102**	0.00160***	0.000644	0.000978**	0.00155***	0.000541	0.000792	0.00146***	0.000417
	(0.000448)	(0.000467)	(0.000666)	(0.000465)	(0.000431)	(0.000672)	(0.000523)	(0.000463)	(0.000681)
R-squared	0.143	0.142	0.086	0.203	0.211	0.135	0.244	0.270	0.181

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Notes: The sample consists of college graduates who were 18 between 2003 and 2010 who graduated between 2010 and 2015. The dependent variable is the maths intensity index of the course of study. Each regression includes the survey year, year of graduation and municipality of residence fixed effects.

Chapter 3

Career Effects of Educational Mismatch

3.1 Introduction

There is increasing evidence that labour market conditions faced by young workers upon completing education can have substantial long term effects on their careers. The issue is of increasing relevance in light of the latest economic crisis, that made young workers face severely adverse macroeconomic conditions upon labour market entry. Previous research has shown that even short-lived labour market shocks can cause persistent career losses, suggesting that the recent cohorts of graduates might still be bearing the costs of the Great Recession.[1]

Exploring the mechanisms through which recessions affect labour market outcomes is crucial to help shaping policies in support of the negatively affected cohorts.

Recessions are associated with an overall lower supply of jobs; thus, young graduates can take more time to find their first job and spend more time in unemployment. Unemployment can have long-term effects on future labour market possibilities, because of depreciation of human capital (Becker, 1994), psychological discouragement or habituation effects (Clark et al., 2001), or employers using individuals' unemployment as a signal of low productivity (Lockwood, 1991).

[1]See, for example, Oyer (2006) for an analysis of MBA college graduates and PhD economists; Genda et al. (2010) for a comparison of US and Japanese college graduates; Oreopoulos et al. (2012), Altonji et al. (2016) and Liu et al. (2016) for the analysis of respectively Canadian, US and Norwegian college graduates.

During recessions the wage offer distribution worsens and the literature has also documented that the quality of jobs tends to decline, for example in terms of opportunities for promotions and training (Gibbons and Waldman, 2006). Hence, a poor early start could put workers in low paying/lower quality jobs. The initial shocks can become persistent because of contracting rigidities and search frictions that make it difficult to move to a better job. Oreopoulos et al. (2012) provide evidence that the career losses of Canadian college graduates entering the labour market during a recession are explained by the low quality of the first employer – measured in terms of firm size and average earnings among employees. A similar result is obtained by Oyer (2006), who finds that PhD economists graduating during a recession start in lower ranked universities, and they do not move to better institutions because they are less productive in terms of publications.

Furthermore, there is evidence that idiosyncratic match quality is affected by the tightness of the labour market (Hagedorn and Manovskii, 2013; Frühwirth-Schnatter et al., 2010). The evidence provided by Liu et al. (2016) suggests that the main mechanism behind the negative effect of poor labour market conditions upon entry on earnings is the quality of the first job in terms of match between skills of workers and skills requirements of industries in which these workers are employed, and they show that the effect is reduced when workers are able to switch to the right industry.

In this study I analyse the short and medium term career outcomes of Italian individuals completing high school or university between 1993 and 2010, as a function of the macroeconomic conditions upon entering the labour market after completion of education. I combine individual data from the Italian Labour Force Survey with a unique dataset on job vacancies advertised on Italian newspapers to analyse how workers with different skills react to higher competition upon entry in the labour market. The labour market relevant for each individual is defined by the level of education possessed, the geographical area and year of entry. The degree of competition is measured with the ratio between the number of unemployed workers possessing a given level of education and the number of job vacancies requiring the same level of education, in a given region and year. This measure indicates the extent of the discrepancy between the labour supply and labour demand according to the dimension of education, and I indicate it as "educational mismatch".

I investigate the following mechanism: Barnichon and Zylberberg (2018) show that during a recession some high-skill workers try to escape competition from their high-skill peers by moving down the occupational ladder.[2] I analyse whether workers who face very high competition upon entry in the labour market compete with workers with lower level of education, ending up overeducated and suffering wage losses.[3]

My paper contributes to the recent literature trying to assess the long term impact of entering the labour market in bad times in at least three ways. First, I emphasise the role of the labour demand dimension by measuring labour market conditions at entry in terms of not only the unemployment rate but also looking at the characteristics of jobs offered. Second, I highlight the aspects of heterogeneity of the effect and of competition across workers with different skills. Finally, I provide evidence for Italy, a country with very persistent unemployment rate and high youth unemployment rate, and with a high incidence of the overeducation phenomenon.

I find that both college and high school graduates who enter the labour market when "educational mismatch" is higher have lower probabilities of being employed and lower salaries, even many years after graduation. Moreover, these workers are more likely to be employed in a job for which a lower level of education is required. The evidence suggests that workers with different skills compete with each other, and in particular some workers try to escape strong competition from their same-skill peers by taking jobs for which a lower level of education is required.

The rest of the chapter is organised as follows: Section 3.2 presents some features characterising the Italian labour market and describes the data used; Section 3.3 discusses the identification strategy used for the empirical analysis. Results are presented and discussed in Section 3.4. Section 3.5 summarises the conclusions.

[2] The underlying assumption is that when workers with different skills compete for the same vacancy, high-skill applicants are systematically hired over less-skilled competing applicants. In Barnichon and Zylberberg (2018) this "ranking" mechanism is generated endogenously by the model through the wage bargaining process.

[3] It is a well-established result in the empirical literature that overeducated workers get lower wages than workers with similar education but in jobs in which their schooling equals what is required. See Leuven and Oosterbeek (2011) for an extensive survey on the economics literature on overeducation.

3.2 Background and Data

3.2.1 Demand and Supply of Skills in Italy

Recessions are characterised by a rising number of unemployed job seekers and falling job opportunities, both contributing to a higher unemployment-to-vacancy ratio. These changes might not be uniform across the skills distribution.

Sahin et al. (2014) show that in the latest recession the rise in unemployment was partly due to the mismatch of unemployed workers and vacancies across distinct labour markets, defined by industries, occupations or geographical locations. They compare the actual allocation of unemployed workers across sectors to an ideal allocation that would be selected by a planner who faces no impediment in moving idle labour across sectors. They show that this optimal allocation comes from equating efficiency-weighted vacancy-unemployment ratios across sectors and use this optimality condition to construct a "mismatch index"; this index is between 0 and 1 and gives a measure of the fraction of hires lost every period because of job seekers misallocation. Following their work, I compute an index of mismatch of vacancies and unemployed workers across distinct labour markets defined by the level of education – less than high school degree, high school diploma and college degree – and the geographical macroarea – north east, north west, central Italy, southern Italy, islands – in the years between 1993 and 2010.[4] The time series of this mismatch index, which is illustrated in Figure 3.1, indicates that between 11 and 24% of hires were lost each year in the period considered because of the misallocation of job seekers across education groups and geographical location. The resulting index for Italy is higher than the index obtained by Sahin et al. for US across occupations and geographical location[5], suggesting that the phenomenon of mismatch between demand and supply across the education and geographical location is a severe issue in Italy.

One implication of this labour market feature can be that there is competition across skills and that a phenomenon of "trickle down unemployment" takes place:

[4] I apply Sahin et al. (2014) formula for the mismatch index in the simplest case of absence of heterogeneity with respect to matching efficiency, productivity and job destruction, and using as vacancy share $\alpha = 0.5$. The formula is $M_t = 1 - \sum_i^I \sum_j^J (\frac{v_{ijt}}{v_t})^\alpha (\frac{u_{ijt}}{u_t})^{(1-\alpha)}$.

[5] See Figure B15 of Sahin et al. (2014) for the plot of the mismatch index across 19 2-digits occupations and 9 US census divisions obtained for US for the period May 2005-June 2011, that ranges from approximately 0.11 to approximately 0.16.

some workers escape competition from their same skill peers by competing with lower skilled workers; workers take jobs requiring a lower level of education than the one they possess, ending up overeducated, and moving down the occupational ladder.

The 2014 report on the Italian labour market from the Italian National Institute of Statistics (ISTAT) presents some interesting evidence in this direction. In its chapter discussing the situation of the Italian labour market during the years of the Great Recession, the report has a focus on the characteristics of the labour force by level of education. By distinguishing workers with high (at least bachelor degree), medium (lower secondary and secondary education) and low (primary school) level of education, the report highlights three facts characterising the period 2008-2013: (i) the increase of the active population with high level of education as opposed to the decrease in the group with a low level of education, indicating a shift in the supply of labour towards more skilled labour; (ii) the number of workers with high level of education who are employed increased, but the increase was mainly driven by employment in jobs requiring medium and low skills; (iii) the number of workers with low level of education who are employed decreased. This evidence is interpreted as suggestive of the fact that highly educated workers might have been able to protect themselves from the crisis by taking jobs requiring less than the level of education possessed, pushing workers with lower level of education out of employment.

Two additional pieces of evidence point to the same direction and support this idea. The first one is that the unemployment rate of high skill workers appears to be less cyclical relative to the one of lower skilled workers. Figure 3.2 plots the unemployment rate in Italy in the last twenty years by level of education and shows that over the period considered the unemployment rate of highly educated workers has the lowest negative correlation with the GDP growth.[6] This is suggesting that Italian high skilled workers are more sheltered from macroeconomic conditions with respect to their low skilled counterparts.

The second piece of evidence comes from the European Community Household Panel data, from which it is possible to obtain a measure of the overeducation phenomenon across European countries. The survey asks the questions "*Do you feel that*

[6]The correlations between GDP growth and unemployment rate growth for primary, secondary and tertiary education are respectively -0.64, -0.71 and -0.46.

you have skills or qualification to do a more demanding job than the one you have now?" and "*Have you had formal training or education that has given you skills needed for the current type of work?*"; workers who answer yes to the first question and no to the second one are classified as overeducated. Figure 3.3 plots the percentage of overeducated workers averaged across the period 1994-2001 for some European countries and shows that in Italy the overeducation incidence is 26%, the highest among the countries considered.

3.2.2 Data on Unemployment and Job Vacancies

In order to obtain a measure of the mismatch of demand and supply of labour in a given labour market defined by the level of education and the geographical location, I combine two data sources: individual microdata from the Italian Labour Force Survey (LFS) – collected and made available by the ISTAT – and microdata on job vacancies from the Help Wanted Time Series dataset made available by ISFOL – *Istituto per lo Sviluppo della Formazione Professionale dei Lavoratori*.

Following Destefanis and Fonseca (2007), I use individual data from the ISTAT LFS to measure stocks of unemployed workers and labour force participants by geographical area and degree of education possessed. I compute the number of unemployed workers by year (from 1993 to 2010), five macroareas (north west, north east, central Italy, southern Italy, islands), and three levels of education (less than high school, high school degree and college degree).[7] The stock of unemployed workers is measured in the second quarter of each year.

To measure the labour demand side, I exploit a unique micro-level dataset of vacancies, the ISFOL-Help Wanted Time Series. It collects the vacancy advertisements published on the main Italian newspapers, which have been classified and filled throughout the years by the *Centro di Statistica Aziendale* (CSA) in Florence on behalf of ISFOL.[8]

The collection of advertisements by CSA begun more than 30 years ago and

[7]I will include in the "high school" category only the unemployed workers who obtained a high school degree that allows access to university; in Italy there is also the possibility for some secondary education institutes (specifically the ones with *technical* or *professional* curriculum) to obtain a 3 or 4 years qualification (*qualifica*), that does not allow access to university afterwards: I will include students with this type of qualification in the "less than high school" category.

[8]I am grateful to Michele Cuppone for making the data available.

it continued until 2010. It started with the registration of the job offers from the advertisement sections of the most important national newspapers; throughout the years it expanded both in terms of newspapers covered and in terms of the information collected for each job advert.[9] These data represent a *unicum* in Italy and also at the European level. Its publication has been followed by a discrete number of studies that have used them as valid proxy of the labour demand in Italy, in particular to study the relationship between vacancies and unemployment known as the Beveridge curve (see for example Mocavini and Paliotta Isfol (2005), Destefanis and Fonseca (2007)).

The database is unique because it allows to work with vacancies at a micro rather than aggregate level: each observation is a single advertisement. Figure 3.4 provides an example of how an advert is coded into the dataset. From each advertisement present on the newspaper every information is extrapolated, such that in the final dataset for each observation (job advert) there are variables on: date in which it was posted, name of the newspaper, number of job openings per advert, and (if specified) geographical location of work, level of education required, sector of the firm seeking workers, description of the occupation, age range, required work experience and required knowledge of foreign languages. To construct my measure of labour demand, I sum the number of job openings by quarter-year of posting, macroarea of work and level of education required; in the specific example of the figure I will have 10 vacancies in quarter 2 of 2013, in the north-east of Italy, requiring high school education.

A comparable time series in terms of quantity and quality of variables is available for the years from 1993 to 2010. The survey covers all the main national Italian newspapers with their local editions, plus some local and few foreign newspapers. Throughout the years some newspapers have been added to the survey. To ensure that the variation in the number of vacancies from one year to another reflects fluctuations in labour demand rather than a change in the composition of the sample, I restrict the sample to newspapers that are present in the survey in all years. Eight existing newspapers that were added to the survey between 2001 and 2002 are excluded for

[9] See Mocavini and Paliotta Isfol (2005) for an extensive survey on the measurement of job vacancies in Italy.

the vacancies computation. Table A1 in the appendix reports a detailed list of the newspapers covered by the sample with their classification as national or local.

Given the nature of this dataset, there may be issues about potential duplication of a job advertisement in different newspapers, in the same day or after some days, or in the same newspapers some days later, which could lead to count more than once the same vacancy. The advertisements that are present in the same newspaper or in a different newspaper in a window of one month since the first time it has appeared are dropped from the sample.[10]

For each advertisement, if indicated, the area of job is reported at the level of region and macroarea. Approximately 2% of the total advertisements refers to jobs located outside Italy, which are dropped from the sample. For another 15% of job advertisements the detail of the location within Italy is not specified. A small percentage of these observations (30%) belongs to newspapers in which between 80 and 90% of the published ads refers to vacancies in a single macroarea, which is the area where they have their headquarters and the highest diffusion among readers.[11] For these observations, I impute the macroarea of work according to the area of coverage of the newspaper. The other 70% of the observations with missing location refers to jobs advertised in two newspapers that do not have a clear geographical diffusion (namely *Corriere della Sera* and *La Repubblica*); these observations (8% of the total number of advertisements and 12% of the total number of vacancies) are dropped.

The firm seeking workers can indicate the level of education required for the job. Only for 42% of the total ads is a precise level of education required indicated, that I classify in three macro categories: college degree, high school degree, or less than high school degree.[12] If more than one type or level of education is indicated I classify

[10]Following Mocavini and Paliotta (Isfol, 2005) I consider as duplicates the job adverts that have specified the same: profession, number of vacancies available, age range, level of education, geogaphical location of the job, sector of the firm. See (Isfol, 2005) for a discussion of the incidence of the duplication phenomenon in the ISFOL data.

[11]I look at data on newspapers circulation by geographical area from the independent agency ADS (*Accertamenti Diffusione Stampa*). I thank Simonetta Zambelli of ADS for kindly providing me with these data.

[12]Unlike the data on unemployed workers from ISTAT, for the high school diploma, I cannot distinguish between job vacancies requiring the full high school diploma (giving access to the university) or only the 3 or 4 years qualification in the case of high school degree from *Istituto Tecnico* or *Istituto Professionale*. This means I might be overestimating the vacancies in the high school category and underestimating the ones in the less than high school category.

the vacancy as requiring the minimum level of education among the ones reported. For the observations with non specified level of education required, I impute the level of education most frequently indicated for jobs referring to the same profession, if at least 60% of the overall adverts for the same profession have non missing education.

Figures 3.5 and 3.6 show, respectively, the time series of the unemployment rate obtained from ISTAT LFS and the number of vacancies calculated from ISFOL-HWTS (normalised by the number of employed workers in the same quarter of the previous year), by macroarea and level of education. Southern Italy is characterised by overall higher unemployment rates and much lower level of vacancies for all levels of educations. In all areas the unemployment rate is lower for the tertiary education level, while the labour demand is higher for this group of workers.

Quality of Vacancy Data

Given the nature of the data used to measure labour demand, there may be concerns about their quality: one may wonder whether job vacancies posted on newspapers are a good measure of labour demand.

One concern is that there is a progressive decline in the use of newspapers as a job advertising tool throughout the period considered, so that the variation in vacancies reflects this rather than the actual change in labour demand. A decreasing use of newspaper recruitment can translate in a downward trend that could artificially exacerbate the drop in job vacancies of last recession.

By looking at some statistics about the job search process in Italy over last 20 years, reported in table A2 in the appendix, we see evidence that the importance of newspapers as job search channel has not decreased significantly in the period considered, despite the increased use of other sources as internet. In the top panel I report the distribution of employed workers according to the channel through which they found their job, obtained from different surveys in different years. The evidence suggests that the job search process in Italy is mostly informal, with the channel of relatives and friends being the most relevant. For the overall formal process, newspapers advertisements are the most important channel, and its importance does not seem to decline over time. Moreover, the informal channel includes the use of newspapers, as it is possible that friends or relatives refer about jobs of which they knew through a newspaper. In the bottom panel, I report the percentage of job seekers who declare they examined job advertisements on newspapers and on the internet as part of their search activity, by geographical macroarea. The use of newspapers advertisements as job search tool is higher in northern regions; in all macroareas the probability of having examined newspapers advertisements shows a decline over time, but this is negligible and more importantly it does not seem to be dramatically different across areas.

To address the issue of the quality of the vacancy data, I investigate whether they are good predictors of job finding. In particular, I analyse the relationship between the measure of vacancies in different years, geographical locations and levels of education obtained by ISFOL and the distribution of hires along the same dimensions obtained from the ISTAT LFS.

Following the large literature that estimates a matching function using aggregate data on unemployment and vacancies[13], I estimate the following model:

$$\begin{aligned} ln(Hires)_{qme} = \alpha + \beta_1 ln(Vacancies)_{qme} + \beta_2 ln(UnemploymentInflow)_{qme} + \\ \beta_3 ln(UnemploymentStock)_{(q-1)me} + \gamma_q + \delta_m + \epsilon_{qme} \end{aligned} \tag{3.1}$$

where the number of hires in quarter q, macroarea m and level of education e is regressed on the flow of vacancies posted each quarter, the inflow of unemployed workers in the same period and the stock of unemployed workers in the previous quarter – all measured along the same dimensions – plus quarter and macroarea fixed effects. Hires are obtained from ISTAT LFS as the number of people who declare they started their job in the year and quarter of the interview; unfortunately I am not able to disentangle the flow into employment from unemployment from the flow into employment from other employment. The inflow of unemployed workers is obtained as number of unemployed workers who declare they stopped working the same year and quarter of the interview; in this case, I am only able to catch the flow into unemployment from employment.

[13]See Petrongolo and Pissarides (2001) for an extensive survey.

Table 3.1 reports the results of the estimation of equation [3.1] both when not including the stock of unemployed workers in the previous quarter – columns (1) and (2) – and when including it, in both cases respectively with and without macroarea fixed effects. The coefficient of the number of vacancies is always significant at least at the 5% confidence level. The results from all different specifications show that the ISFOL vacancy data are effective in predicting job finding, and provide evidence in support of considering ISFOL data a good measure of labour demand.

3.2.3 Individual data on Labour Market Outcomes

The three main outcomes of my analysis are the employment status, monthly earnings and the probability of being overeducated. The outcomes are measured from individual cross-sectional quarterly data on Italian workers – the ISTAT Labour Force Surey – which contain information on respondents' highest educational attainment and on the year in which they completed education. I focus on individuals with high school degree, or with college degree and above (from undergraduate to master's degree and PhD, including degrees issued by art institutions that in Italy are recognised as university degree). Employment status and overeducation are measured from 2005 to 2010.[14] Worker's monthly salary is instead observed only in 2009 and 2010, because ISTAT started collecting data on earnings in 2009.

I construct the overeducation variable from the distribution of the education level possessed by workers aged 18 to 35 employed in each of the 121 3-digits occupations from the ISTAT 2001 classification of professions (I focus on the distribution observed in the 2005 wave of the survey in order to reduce the risk that my measure is influenced by the business cycle). The 121 3-digits ISTAT professions are a detailed classification within nine broad groups: 1 - Legislators, entrepreneurs and managers; 2 - Professionals; 3 - Technicians; 4 - Clerks; 5 - Service workers and shop and market sales workers; 6 - Craft and related trades workers and agricultural workers; 7 - Plant and machine operators and assemblers; 8 - Elementary occupations; 9 - Armed forces. Each worker in my sample will be classified as overeducated if the majority of individuals employed in his same profession has a lower level of education than the

[14] I exclude the 2008 wave because in this year the information on the year of graduation is missing.

one he possesses. As a result, college graduates employed in low-skills jobs (groups from 4 to 9) and middle-skills technical jobs (group 3), plus entrepreneurs, some managers and some professionals, are considered overeducated. In some robustness checks I will adopt more conservative measures, namely considering as overeducated: (i) workers who are employed in professions for which at least 60% of individuals have a lower level of education; (ii) college graduates employed in low-skills jobs only (groups 4 to 9) and high school graduates employed in elementary occupations only (group 8).

The sample of individuals in the labour force observed from 2005 to 2010, whose highest educational attainment is at least high school, is composed of approximately 670,000 individuals, of which around 200,000 attained college education. The dataset provides the information on the year in which the highest degree was obtained. When the information is missing, workers are asked to report the age at which they completed their education. This happens only for 8% of college graduates, but for 26% of high school graduates. I focus on the sub-samples of individuals with non-missing year of education. For college graduates, the distribution of individuals across years, age and macroarea of graduation is not different from the sample with non-missing year of graduation, and the results are unchanged when these observations are included in the analysis. For high school graduates, the distribution of age at graduation is significantly different; in particular it presents an abnormal spike at the age of 19 (which is the most common age at which individuals complete high school in Italy), which I attribute to a possible recall error. For this reason, I prefer excluding this group of observations; in the Appendix, Figure A1 illustrates the issue in greater detail and table A8 presents results from the analysis performed on the full sample.

I then focus on college graduates aged between 22 and 32 at graduation and high school graduates who completed high school between the age of 18 to 21, both not enrolled in education in the years in which the outcome is observed (excluding in this way 16% of the college graduates sample and 17% of the high school graduates sample). Since I have data on macroeconomic conditions from 1993 to 2010, I then restrict the samples to individuals entering the labour market in those years. The final samples are made by 78.892 college graduates and 99.200 high school graduates.

Table 3.2 shows some summary statistics. There are 17 cohorts of college and high school graduates. Workers with high school degree earn on average more than 300 Euro less per month than their higher skilled counterpart; individuals with college degree have on average higher probability of being employed but also higher probability of being overeducated. A relatively high percentage of college graduates (slightly more than half) is classified as overeducated when looking at the mode of education level by profession; with the alternative measure that considers as overeducated workers with college degree employed in low-skills professions this percentage goes down to 20%.

3.3 Empirical Strategy

The goal of this study is to estimate the effect of macroeconomic conditions upon entry in the labour market – measured in terms of mismatch between demand and supply of labour in the labour market relevant for each worker – on young workers' careers. The relevant labour market for each individual is defined by his level of education and the year and geographical area of exit from education.

The cross-sectional nature of the data used does not allow to observe the geographical location at time of graduation: as proxy for macroarea at entry, I use the current macroarea of residence. This introduces potential attenuation bias due to measurement error. Namely, if young workers move after graduation to an area where the labour market is in better shape, I am underestimating the effect of labour market conditions upon entry. Thus, if I find an effect, this must be a lower bound. The measurement error due to migration is attenuated by the choice of the geographical unit: even if workers move across regions and municipalities after graduation, they are less likely to move across macroareas. Moreover, in general mobility rates within Italy are very low. In my sample of high school and college graduates the percentage of workers who declare to have moved from another region to start the current job is less than 2% in all years, and among this minority still some of them could have moved within the same macroarea. Some evidence on migration patterns of college graduates comes from the AlmaLaurea survey on Graduates' Employment Conditions. Table A3 reports the distribution of workers with a college degree by area of work for each area of study. More than 90% of individuals graduating in north west stay in the same area to work, and this proportion is approximately 80% for people graduating in north east and central Italy. Migration from the south and the islands is higher, but still more than 70% of students graduating in these areas do not move to work.

I estimate for both samples of high school and college graduates the following specification:

$$\begin{aligned} y_{icrt} = \alpha + \eta X_{icrt} + \beta_1 ln(U/V)_{cr} + \beta_2 ln(U/V)_{cr} * YSG + \\ \gamma_1 ln(U/V)_{rt} + \gamma_2 ln(U/V)_{rt} * YSG + \phi_t + \delta_c + \theta_r + \epsilon_{icrt} \end{aligned} \tag{3.2}$$

The outcome variable y is employment status, salary or probability of being overeducated, measured in the year of interview t for an individual i with a given level of education, graduated in year c, who lives in region r at the time of the survey. The main independent variable is the ratio between number of unemployed workers with the same education level of the individual and number of vacancies requiring the same level of education, measured at time of graduation c in region r; this term is also interacted with the number of years since graduation to investigate how persistent the effect is. In order to estimate the isolated temporary shock of initial labour market conditions holding everything else constant, the model also controls for labour market conditions at the time of interview t, that are interacted with years since graduation to allow the effect to be different according to which stage of the career the worker is in. As a robustness check, I estimate an alternative specification where the U/V ratios at time of interview are not included, and a full set of interactions between year of interview and macroarea dummies is included. X is a set of individual control variables, including age, gender, marital status and number of years since first entry in the labour market. Year of interview, cohort and macroarea fixed effects are included in the model. To account for group-specific error components, standard errors are clustered at the cohort-region level.

Hence, the main coefficient of interest β_1 measures the impact of labour market conditions upon graduation over and above impact of current conditions. The coefficient β_2 measures how the impact of labour market conditions varies with the number of years since graduation, to check for the persistence of the effect and disentangle the short and longer term effect.

Since the Italian labour market is characterised by a high persistence of regional labour market conditions differentials, one may lack statistical power for identifying the effect of interest, having controlled for both year and macroarea fixed effects. Table A4 in the appendix shows the very high correlation of the U/V ratios time series across macroareas. Table A5 provides evidence showing that the fraction of the variation in the independent variable that is due to between-groups (macroarea) variation is only between 47 and 52%. For this reason I will show the results for each outcome both when macroarea fixed effects are included and when they are not.

3.3.1 Identification issues and challenges

The empirical strategy relies on the key identification assumption of year and macroarea of graduation being *as good as random*, conditional on all else controlled for. This ensures that the variation in the U/V ratios arises from changes in aggregate labour demand and supply that are uncorrelated with characteristics of different graduation cohorts. This is not the case if individuals endogenously choose when and where to graduate.

For example individuals may choose to obtain their education in a region where macroeconomic conditions are better relative to where they were born. Students might also postpone their exit from education if they perceive the labour market is extremely slack at the time they are supposed to graduate. Both issues are less severe for high school graduates: they enrol at the age of 14 usually in a high school in proximity of their area of residence. High school students progress automatically to the following grade unless severe failure in meeting some performance requirements; failure of being admitted to the following year is perceived as very negative and it is harder to think that students would fail on purpose to postpone graduation. High school in Italy lasts 5 years so that the age of exit is between 18 and 19 (depending on the month of birth): 75% of the sample of high school graduates from the ISTAT-LFS 2005-2010 exits high school at age 18 or 19. On the other hand, a bigger concern is that college graduates choose where and when to graduate in response to macroeconomic conditions in the area of residence at the time they are supposed to complete education. In order to account for the potential endogeneity of the choice of the timing of graduation, in a robustness check I impute to all the individuals in the sample of college graduates the year in which they were supposed to exit from university, based on the predicted duration of the study career path. Results of this analysis are reported in last part of the results section.

Educational choice itself may be endogenously related to labour market conditions. Theoretically, two scenarios are possible: individuals decide to undertake higher education when they perceive the labour market for their education level is very slack because the outside option is worse; it is also true that in periods of recession the incidence of the cost of higher education is bigger, and this might induce the less motivated individuals (or the poorest ones) away from pursuing higher education. Results from the empirical literature go in the first direction (see Dellas and Koubi (2003) or Di Pietro (2006)).

I estimate the effect of macroeconomic conditions around time of exit from high school on the probability of being enrolled in university. I look at high school graduates aged 19 to 22 in the survey years, and who were 19 in the years from 1993 to 2010. I estimate the coefficient of macroeconomic conditions at age 19 on the probability of being enrolled in university up to 3 years after high school graduation. Table 3.3 shows the results of this analysis: I indeed find that high school graduates are more likely to be enrolled in university if the macroeconomic conditions in their education cell at time of high school graduation are worse, while a slacker labour market for college graduates makes high school graduates less likely to enrol in university. Hence, during a recession more individuals stay in education after high school, and one may expect composition effects of such decision, i.e., (i) the average quality of high school graduates entering the labour market may be lower and (ii) the average quality of university enrolment may also be lower such that the college graduates I observe later are negatively selected.

On one end, these selection effects may lead to overestimate the negative impact, if any, of labour market conditions on the careers of high-school graduates. On the other hand, one would expect the opposite bias in the estimated effect of labour market conditions on the careers of college graduates. These potential biases need to be kept in mind while interpreting my estimated effects of interest.

3.4 Results

I estimate equation [3.2] separately for the two samples of college graduates and high school graduates, for each of the three outcomes: (i) the probability of being employed; (ii) monthly salary; (iii) the probability of being overeducated.

3.4.1 Outcomes for College Graduates

Table 3.4 reports the results for the sample of college graduates.

The first three columns show that the effect of worse macroeconomic conditions at graduation – measured as ratio between the number of unemployed workers with college degree and the number of vacancies requiring college degree in the area and year of graduation – on the probability of being employed upon entry in the labour market is negative and significant at the 1% confidence level. The result is robust to the inclusion of macroarea fixed effects – column (2) – and to an alternative specification where I do not control for contemporaneous macroeconomic conditions but I include a full set of interactions of year and macroarea dummies – column (3) –, although the coefficient is decreased in magnitude in both cases.

For ease of interpretation, I will calculate the impact of an increase of 50% in the U/V ratio, typical in a recession. Following such an increase the employment probability upon graduation falls by between 2.2 and 5.3 percentage points, which is between 4 and 10% of one standard deviation from the average probability of being employed of college graduates upon exit from college. The interaction with the potential years of experience is positive and statistically significant, indicating that the effect fades over time, and it disappears within between 5 and 9 years from graduation.

Columns (4) to (6) and (7) to (9) show the results respectively on the monthly salary and on the probability of being employed in a profession for which the majority of workers have a lower level of education. Having found a negative effect on the employment probability, in interpreting these results it has to be considered that the effect is estimated on a selected sample of employed individuals. Moreover, data on salaries are only available for a subset of years, namely 2009 and 2010, which explains the smaller sample size. An increase of 50% in the U/V ratio is associated to a decrease in salary of 2.3% upon graduation (column (4)) which persists many years later, and to an increase in the probability of being overeducated upon entry in the labour market of 1.4 percentage points – that is approximately 3% of one standard deviation from the average probability of being overeducated of college graduates upon exit from college – which fades approximately 7 years after graduation (column (7)). These results however are not robust to the inclusion of macroarea dummies, nor to the inclusion of time-area interactions.

The results on the probability of being overeducated are robust to the adoption of the two alternative, more conservative, measures of the outcome, for which the estimations are presented in table A6 in the appendix. The coefficients of the effect of the macroeconomic conditions upon graduation are even higher, especially when considering only college graduates working in low-skills jobs (column (4) to (6)) – for example college graduates working as cashiers or sales assistants. The evidence on the overeducation variable is suggestive of the fact that high skilled workers try to escape higher competition by taking jobs for which a lower level of education is required.

In order to account for potential endogeneity of the choice of the timing of college graduation, I impute to college graduates a predicted year of graduation. Since I do not have information on the legal duration of the degree course from which each individual graduated, I look at the distribution of age at graduation of individuals observed in 2005 within different levels of tertiary education that I can distinguish[15], and take the mode as theoretical year of graduation for all individuals in the same degree type group. Hence, I estimate the reduced form model where I analyse the effect on the three outcomes of the macroeconomic conditions measured at the predicted year of entry in the labour market. Table 3.5 reports the result of this analysis. The results on the employment status are robust although decreased

[15]Degrees issued by art institutions; university level short-cycle degrees issued by institutions different from universities; undergraduate degrees; master's degrees; single-cycle master's degrees; post-graduate specialisations;

in magnitude, while for the other two outcomes the effect is less precisely estimated.

3.4.2 Effects for High School graduates

The findings for college graduates are confirmed when analysing the sample of high school graduates, that are presented in table 3.6: worse macroeconomic conditions upon exit from high school have a negative effect on employment status and salary and a positive effect on the probability of being overeducated, with the effect on unemployment being stronger and more robust relative to the other outcomes. All the considerations about the inclusion of macroarea dummies and the selection of the sample of employed workers made above apply here as well.

Relative to college graduates, the effect on employment status is slightly lower in magnitude but more persistent: an increase of 50% in the ratio between unemployed workers and job vacancies is associated to a decrease in the probability of being unemployed from 1.5 to 4.5 percentage points upon entry in the labour market (columns (1) to (3) of Panel A); the effect fades at a slower rate and it persists up to 16 years after graduation. The effect on salary is smaller and more persistent but less robust to the alternative specifications (columns (1) to (3) of Panel B). Finally, in Panel C we see that high school workers facing higher educational mismatch upon exit from school have higher probability of ending up overeducated – between 0.8 and 3.3 pp higher – and of staying so until at least 10 years after graduation. Table A7 in the appendix shows that the results on overeducation are robust to the alternative measures of the outcome.

In order to investigate whether there is competition across workers with different skills, I estimate model [3.2] adding the labour market conditions upon graduation measured at the higher education cell (college). Results are shown in columns (4) to (6) of each panel respectively for the three different outcomes. The macroeconomic conditions measured at the college cell have a significant effect on all the measured labour market outcomes for high school graduates, indicating that there is competition across skills. The effect on the probability of being employed is negative and slightly lower than the own macroeconomic conditions effect (Panel A). For salary and overeducation probability, once the college U/V ratios are included, the coefficient of the own macroeconomic conditions loses significance. This evidence is suggesting that workers with higher skills facing bad macroeconomic conditions start competing for with lower skilled jobs, moving down the occupational ladder. Higher educational mismatch faced by higher skills workers upon entry in the labour market pushes lower skilled workers into worse jobs, both in terms of salary and of skills-match, or out of employment.

3.5 Concluding Remarks

In Italy labour demand and labour supply are mismatched in terms of level of education possessed by workers and level of education required for the vacant jobs available. Furthermore, the overeducation phenomenon has a relative high incidence when compared to other European countries.

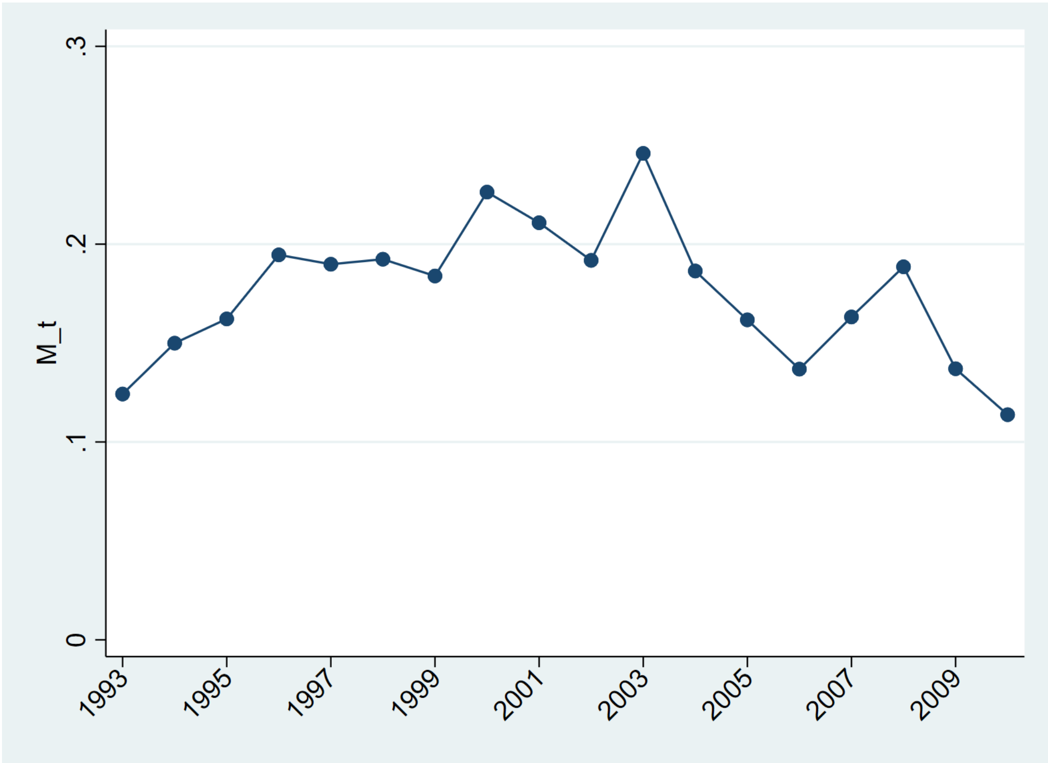
In this paper I analyse the career effects of entering the labour market in periods of bad macroeconomic conditions for Italian young workers graduating from high school or university in the years between 1993 and 2010. I estimate the impact on labour market outcomes in the short and medium run of educational mismatch – namely the degree of discrepancy between the labour supply and labour demand according to the dimension of education possessed by the former and education required by the latter. I study the extent to which workers with different levels of education compete in the labour market, and provide evidence of differential effects across these type of workers.

I find that both college and high school graduates who enter the labour market when educational mismatch is higher are less likely to be employed and have lower salaries even several years after graduation. Differently from the findings of the literature for North America and Northern Europe, I find the strongest effect on employment status, which could reflect the peculiarity of the Italian labour market characterised by rigidity of salaries. Moreover, I find that in a labour market characterised by high educational mismatch, workers have higher probability of ending up in a worse match in terms of skills (as measured by the level of education).

My results suggest that workers who complete education during bad times and enter a labour market characterised by a higher level of educational mismatch compete with workers with lower level of education to escape from competition from their same skill peers. The occupational ladder moves down: higher skilled workers will be overeducated and lower skilled workers will be pushed in jobs requiring an even lower level of education or in unemployment.

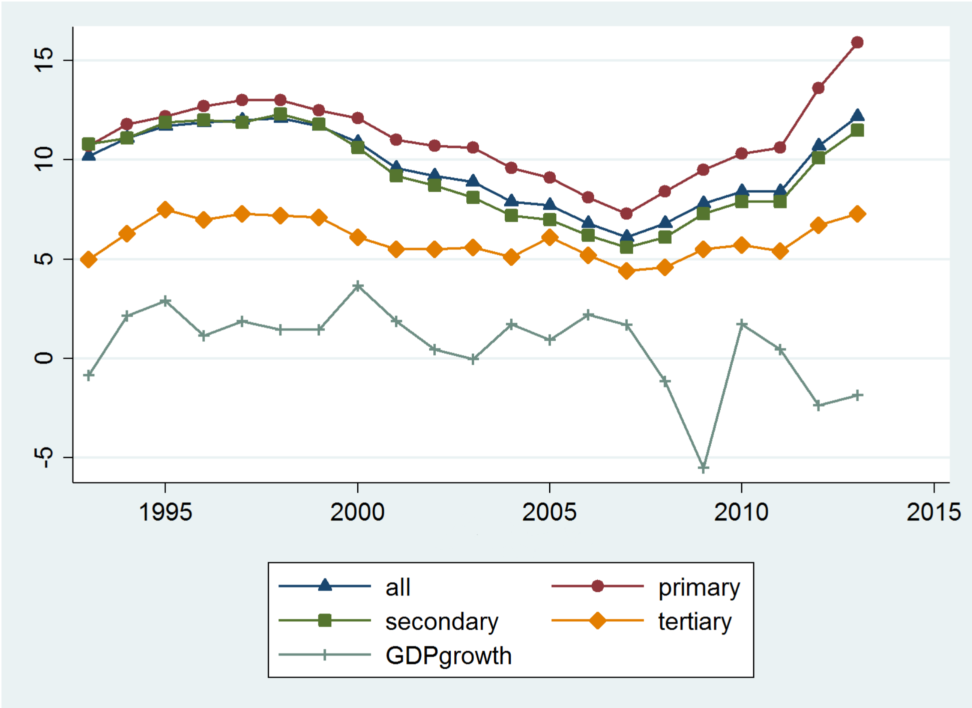
Figures and Tables

Figure 3.1: Mismatch index



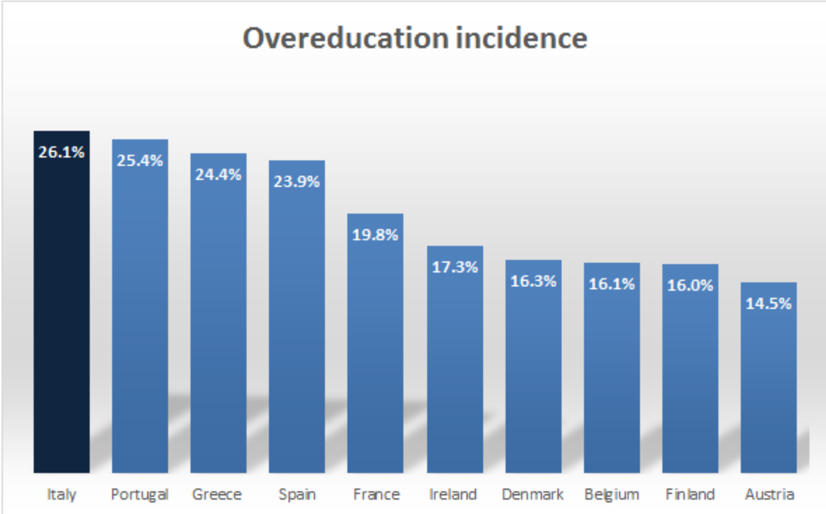
Notes: The figure plots a *mismatch index* across geographical location and educational level for the period 1993-2010. The index is computed using Sahin et al. (2014) formula in the simplest case of absence of heterogeneity with respect to matching efficiency, productivity and job destruction: $M_t = 1 - \sum_i^I \sum_j^J (\frac{v_{ijt}}{v_t})^\alpha (\frac{u_{ijt}}{u_t})^{(1-\alpha)}$ - where i are 3 levels of education and j 5 macroareas - and using as vacancy share $\alpha = 0.5$.

Figure 3.2: Unemployment rates by education groups



Notes: The figure shows the time series of the GDP growth and of the unemployment rate from 1993 to 2013 by level of education (less than primary, primary and lower secondary education; upper secondary and post-secondary non-tertiary education; short-cycle tertiary education, bachelor or equivalent, master or equivalent and doctoral or equivalent).

Figure 3.3: Overeducation incidence



Notes: The figure shows the average percentage of overeducated workers obtained from the European Household Panel Survey 1994-2001 for the European countries for which the information is available in the survey. Workers are classified as overeducated if they answered yes to the question *"Do you feel that you have skills or qualification to do a more demanding job than the one you have now?"* and no to the question *"Have you had formal training or education that has given you skills needed for the current type of work?"*.

Figure 3.4: Example of a newspaper job advertisement

SCODRO S.R.L. CERCA
AGENTI PLURIMANDATARI
Per le province di
TREVISO, BELLUNO, VICENZA, PADOVA, VENEZIA E VERONA

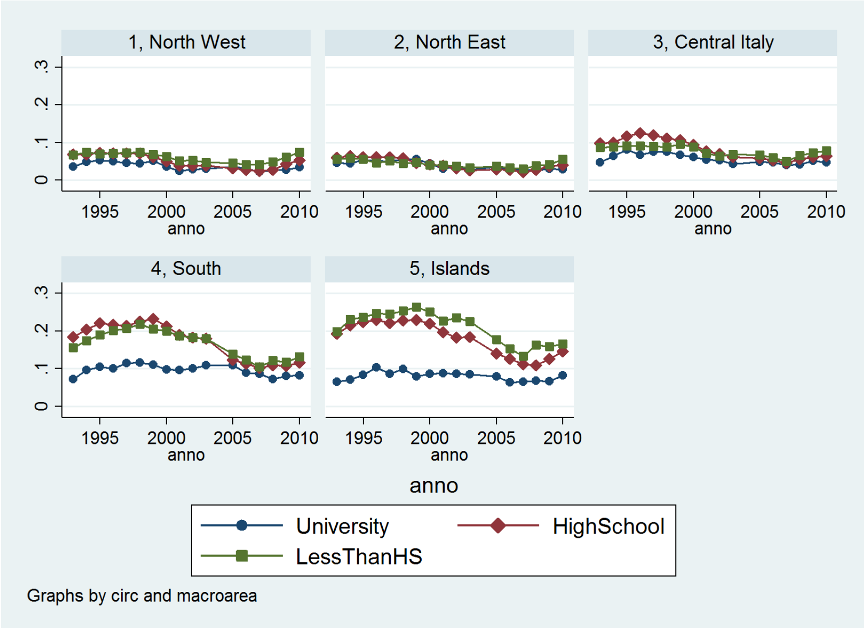
La persona che stiamo cercando, si occuperà di ampliare e sviluppare nuovi settori di mercato e di potenziare quelli già esistenti, coordinandosi con il Responsabile Commerciale. OFFRIAMO sistema di provvigioni in esclusiva [...].CERCHIAMO candidati in possesso di P.IVA, che abbiano già maturato esperienza nel commercio di prodotti legati alla ristorazione, con una forte determinazione e buone doti relazionali. [..]In presente annuncio si rivolge a candidati di ambo i sessi

Data di pubblicazione	**22/05/2014**
Categoria - Ruolo	**Commerciale Vendite**
Settore	**Alimentare**
Titolo di studio richiesto	**Diploma**
Numero posizioni aperte	**10**
Luogo	**Padova, Venezia, Verona, Belluno e altri...**

Codification of ad in the dataset

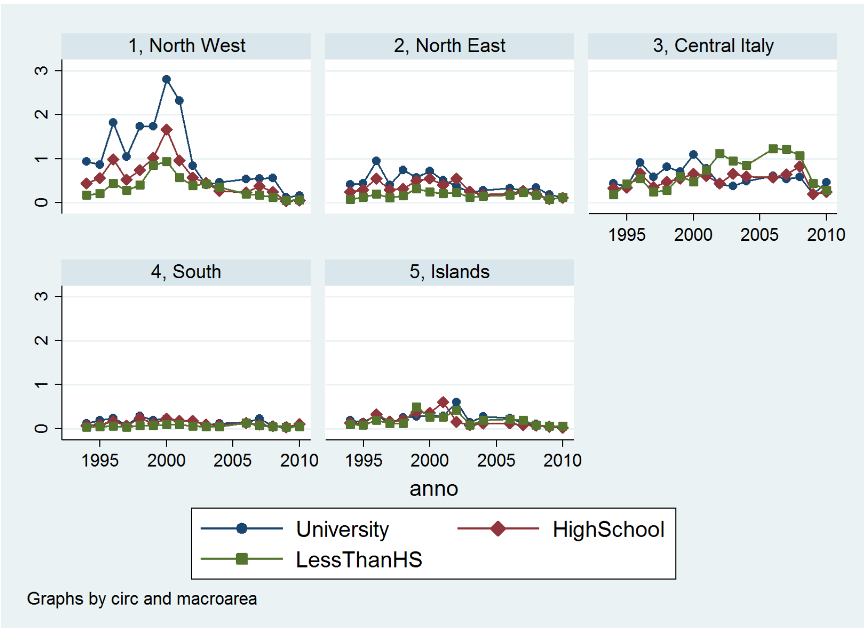
DATA	ANNO	TRIM	GRN
22/05/2013	2013	2	2
REG	CIRC	PROF	QUAL
5	2	AGENTE PLURIMANDATARIO	3342
NMOF	IST1	IST2	ESP
10	DIPLOMA		X
AREA	AZND	SEDLEG	NAZ
C	SCODRO S.R.L.	5	I

Figure 3.5: Unemployment rate by geographical area and level of education



The time series of the unemployment rate is obtained from ISTAT Labour Force Survey waves of 1993 to 2010.

Figure 3.6: Job vacancies by geographical area and level of education



The figure plots the number of job vacancies in each of the five macroareas in the second quarter of the years from 1994 to 2010, normalised by the number of employed workers in the same quarter of the previous year, obtained from ISFOL-HWTS.

Table 3.1: The Determinants of New Hires

	(1)	(2)	(3)	(4)
VARIABLES				
ln(Vacancies)	0.0914***	0.181***	0.152***	0.0603**
	(0.0208)	(0.0318)	(0.0226)	(0.0271)
ln(Unempl. Inflow)	0.716***	0.651***	0.429***	0.137***
	(0.0250)	(0.0263)	(0.0456)	(0.0398)
ln(Unempl. Stock lagged)			0.315***	0.706***
			(0.0425)	(0.0483)
Area FE	No	Yes	No	Yes
Observations	359	359	344	344
R-squared	0.795	0.832	0.814	0.899

Notes: The number of hires, unemployed workers and vacancies are measured quarterly from 2005 to 2010. The dependent variable is the log of new hires, measured as number of employees who started working in the same year and quarter of the interview. Robust standard errors are reported in parentheses (* significant at 10%, ** significant at 5%, *** significant at 1%).

Table 3.2: Summary Statistics

Panel A. College Graduates

Variable	Obs	Mean	Std. Dev.	Min	Max
Age	78,892	33.31	5.04	22	49
Female	78,892	0.57	0.49	0	1
Married	78,892	0.40	0.49	0	1
YSG	78,892	7.1	4.4	0	17
Salary (2009 €)	22,793	1390	503	346	3248
Employed	78,892	0.91	0.29	0	1
Fulltime	71,498	0.86	0.35	0	1
Overeducated	71,498	0.56	0.5	0	1

Panel B. High School Graduates

Variable	Obs	Mean	Std. Dev.	Min	Max
Age	99,200	27.32	4.50	18	38
Female	99,200	0.45	0.5	0	1
Married	99,200	0.22	0.42	0	1
YSG	99,200	8.4	4.48	0	17
Salary (2009 €)	30,742	1060	338	260	2100
Employed	99,200	0.87	0.34	0	1
Fulltime	86,047	0.86	0.35	0	1
Overeducated	86,047	0.32	0.47	0	1

Notes: Panel A: sample consisting of workers with tertiary education, aged between 22 and 32 at graduation, not enrolled in education in the years in which the outcome is observed. Panel B: sample consisting of workers with secondary education, who completed high school between the age of 18 to 22, not enrolled in education in the years in which the outcome is observed.

Table 3.3: Choice of Level of Education

	(1)	(2)	(3)	(4)
VARIABLES				
lnU/V_hs	0.00521*	0.0217***	0.0162***	0.0261***
	(0.00279)	(0.00715)	(0.00604)	(0.00688)
lnU/V_hs*YSG	-0.00106	-0.0120***	-0.00133	-0.0107***
	(0.00129)	(0.00297)	(0.00133)	(0.00305)
lnU/V_univ		-0.0199***		-0.00139
		(0.00750)		(0.00139)
lnU/V_univ*YSG		0.0136***		0.0118***
		(0.00362)		(0.00368)
year dummies	yes	yes	yes	yes
cohort dummies	yes	yes	yes	yes
area dummies			yes	yes
Observations	128,271	128,271	128,271	128,271
R-squared	0.029	0.030	0.030	0.030

Notes: The dependent variable is the probability of being enrolled in university for individuals with high school diploma aged 19 to 22 from 1993 to 2010. Controls include gender, marital status, year and macroarea dummies. Robust standard errors are reported in parentheses (* significant at 10%, ** significant at 5%, *** significant at 1%). Standard errors are clustered at graduation cohort-macroarea level.

Table 3.4: College Graduates

	Employment status			Monthly salary			Overeducation		
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
UV_univ	-0.106***	-0.0726***	-0.0445***	-0.0463***	-0.00631	-0.00442	0.0282***	0.0131	0.0146
	(0.0113)	(0.0104)	(0.00684)	(0.0158)	(0.0135)	(0.00907)	(0.00966)	(0.0133)	(0.00991)
UV_univ*YSG	0.0118***	0.0126***	0.00886***	0.00168	0.00256**	0.00241***	-0.00374***	-0.00377***	-0.00401***
	(0.00129)	(0.00120)	(0.000607)	(0.00157)	(0.00108)	(0.000483)	(0.00122)	(0.00117)	(0.000552)
UV_contemporaneous	yes	yes	no	yes	yes	no	yes	yes	no
macroarea FE	no	yes	no	no	yes	no	no	yes	no
macroarea*year FE	no	no	yes	no	no	yes	no	no	yes
Observations	78,892	78,892	78,892	22,793	22,793	22,793	71,498	71,498	71,498
R-squared	0.140	0.145	0.144	0.154	0.159	0.159	0.046	0.048	0.048

Notes: All U/V ratios are in logs. Each estimation includes controls for: gender, log of age, marital status, and potential years of experience. Robust standard errors are reported in parenthesis (* significant at 10%, ** significant at 5%, *** significant at 1%). Standard errors are clustered at graduation cohort-macroarea level.

Table 3.5: College Graduates: Imputed Year of Graduation

	Employment status			Monthly salary			Overeducation		
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
UV_univ	-0.0729***	-0.0425***	-0.0402***	-0.0230	0.0184	-0.00216	0.00831	0.000923	0.0159*
	(0.00952)	(0.00759)	(0.00675)	(0.0201)	(0.0205)	(0.0145)	(0.00929)	(0.0110)	(0.00872)
UV_univ*YSG	0.00647***	0.00758***	0.00720***	-0.00150	-0.000188	0.00178***	-0.00124	-0.00115	-0.00299***
	(0.000792)	(0.000608)	(0.000397)	(0.00172)	(0.00128)	(0.000591)	(0.000933)	(0.000892)	(0.000527)
UV_contemporaneous	yes	yes	no	yes	yes	no	yes	yes	no
macroarea FE	no	yes	no	no	yes	no	no	yes	no
macroarea*year FE	no	no	yes	no	no	yes	no	no	yes
Observations	83,942	83,942	83,942	25,329	25,329	25,329	76,028	76,028	76,028
R-squared	0.095	0.100	0.101	0.156	0.161	0.161	0.108	0.109	0.109

Notes: All U/V ratios are in logs. Each estimation includes controls for: gender, log of age, marital status, and potential years of experience. * significant at 10%, ** significant at 5%, *** significant at 1%. Standard errors are clustered at graduation cohort-macroarea level.

Table 3.6: High School Graduates

Panel A. Dependent Variable: Employment Status

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
UV_HighSchool	-0.0910***	-0.0482***	-0.0302***	-0.0553**	-0.0396**	-0.00958
	(0.00832)	(0.00808)	(0.00581)	(0.0238)	(0.0151)	(0.0125)
UV_HighSchool*YSG	0.00557***	0.00674***	0.00434***	0.00151	0.00590***	0.00104
	(0.000720)	(0.000678)	(0.000402)	(0.00253)	(0.00220)	(0.00196)
UV_univ				-0.0390*	-0.0289*	-0.0387**
				(0.0199)	(0.0163)	(0.0171)
UV_univ*YSG				0.00502**	0.00216	0.00387*
				(0.00223)	(0.00198)	(0.00213)
UV_contemporaneous	yes	yes	no	yes	yes	no
macroarea FE	no	yes	no	no	yes	no
macroarea*year FE	no	no	yes	no	no	yes
Observations	99,200	99,200	99,200	99,200	99,200	99,200
R-squared	0.129	0.134	0.135	0.130	0.135	0.135

Panel B. Dependent Variable: Monthly Salary

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
UV_HighSchool	-0.0608***	0.00463	-0.00959	0.0536	0.0157	-0.0116
	(0.0113)	(0.0109)	(0.0100)	(0.0358)	(0.0223)	(0.0170)
UV_HighSchool*YSG	-0.000238	0.00108*	0.00215***	-0.0113***	-4.92e-05	0.00315*
	(0.00104)	(0.000608)	(0.000520)	(0.00367)	(0.00264)	(0.00186)
UV_univ				-0.0696**	-0.00443	-0.000918
				(0.0328)	(0.0236)	(0.0214)
UV_univ*YSG				0.00679**	-0.000378	-0.00127
				(0.00341)	(0.00233)	(0.00195)
UV_contemporaneous	yes	yes	no	yes	yes	no
macroarea FE	no	yes	no	no	yes	no
macroarea*year FE	no	no	yes	no	no	yes
Observations	30,742	30,742	30,742	30,742	30,742	30,742
R-squared	0.191	0.203	0.203	0.194	0.203	0.203

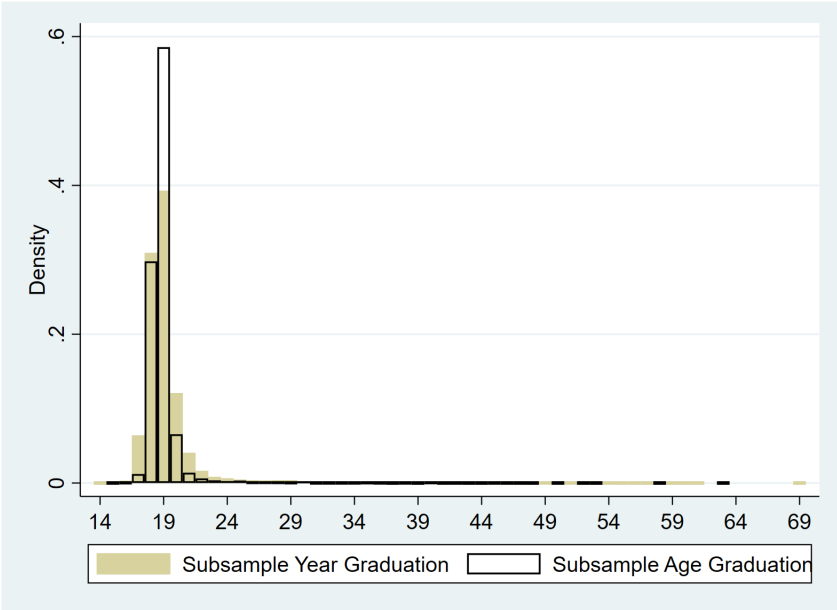
Panel C. Dependent Variable: Overeducation

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
UV_HighSchool	0.0479***	0.0171*	0.00821	0.0188	0.0221	0.00191
	(0.00736)	(0.00943)	(0.00981)	(0.0199)	(0.0141)	(0.0130)
UV_HighSchool*YSG	-0.00159**	-0.00252***	-0.00136***	0.000961	-0.00348*	-0.000223
	(0.000710)	(0.000670)	(0.000467)	(0.00226)	(0.00202)	(0.00161)
UV_univ				0.0539***	0.00882	0.00988
				(0.0190)	(0.0165)	(0.0156)
UV_univ*YSG				-0.00388*	-0.000829	-0.00138
				(0.00220)	(0.00204)	(0.00186)
UV_contemporaneous	yes	yes	no	yes	yes	no
macroarea FE	no	yes	no	no	yes	no
macroarea*year FE	no	no	yes	no	no	yes
Observations	86,047	86,047	86,047	86,047	86,047	86,047
R-squared	0.048	0.050	0.050	0.048	0.051	0.050

Notes: All U/V ratios are in logs. Each estimation includes controls for: gender, log of age, marital status and potential years of experience. Robust standard errors are reported in parenthesis (* significant at 10%, ** significant at 5%, *** significant at 1%). Standard errors are clustered at graduation cohort-macroarea level.

Appendices

Figure A1: Distribution of age upon completion of high school in two sub-samples



Notes: The figure shows the distribution of observations by age at high school graduation for: the sub-sample of individuals who report the exact year of graduation (in yellow); the sub-sample of individuals who report age at graduation from high school (in white). The latter presents an abnormal spike at the age of 19 (which is the most common age at which individuals complete high school in Italy), which I attribute to a possible recall error.

Table A1: Newspapers in ISFOL Sample

Newspaper	**Circulation(Macroarea)**
MESSAGGERO	Italy (headquarters Central Italy)
REPUBBLICA	Italy (headquarters Central Italy)
SOLE 24 ORE	Italy (headquarters North West)
STAMPA	Italy (headquarters North West)
ADIGE	North East
ALTO ADIGE	North East
GAZZETTINO	North East
PICCOLO	North East
RESTO CARLINO	North East - Central Italy
GIORNO	North West
SECOLO XIX	North West
NAZIONE	North West-Central Italy
TIRRENO	Central Italy
TEMPO	Central Italy-South
GAZZETTA MEZZOGIORNO	South
GIORNALE SICILIA	South
MATTINO	South
GAZZETTA SUD	South -Islands
NUOVA SARDEGNA	Islands
SICILIA	Islands
UNIONE SARDA	Islands

Table A2: Job Search Channels in Italy

Panel A Percentage of employed workers by channel through which they found the current job

	ISTAT 2010	Isfol RDL PLUS 2003	Bank of Italy SHIW 1991
Relatives or friends	39.7	27.0	24.5
Self Applications	23.3	22.1	12.1
Help Wanted on the press	13.2	12.6	14.4
Working Experiences	7.1	4.1	n.a.
Training schools and centres	2.1	3.3	2.2
Recruiting personnel agencies	2.4	4.8	n.a.
Internet	0.7	0.0	0.0
Public contests	n.a.	7.5	15.6
Pes	3.0	10.2	31.2
Other	8.4	8.4	0.0

Panel B Percentage of job seekers who have examined newspapers job ads in the reference week (ISTAT LFS)

macroarea	**Overall**	**2005**	**2010**
North West	69	70	67
North East	69	71	70
Central Italy	62	64	61
South	49	50	49
Islands	52	55	46

Table A3: Internal Mobility Rates of Italian College Graduates

Area of graduation	Area of work				
	North West	North East	Central Italy	South	Islands
North West	0.92	0.03	0.02	0.01	0.01
North East	0.11	0.81	0.05	0.03	0.01
Central Italy	0.07	0.04	0.80	0.08	0.02
South	0.12	0.05	0.10	0.72	0.01
Islands	0.13	0.04	0.05	0.06	0.72

Source: AlmaLaurea survey on Graduates' Employment Conditions.

Table A4: Correlations of U/V Ratios

College graduates

	North West	North East	Central Italy	South	Islands
North West	1				
North East	0.8864*	1			
Central Italy	0.5340*	0.4924*	1		
South	0.7818*	0.8336*	0.2058	1	
Islands	0.8759*	0.6959*	0.6098*	0.6227*	1

High School grad.

	North West	North East	Central Italy	South	Islands
North West	1				
North East	0.8658*	1			
Central Italy	0.3587	0.6545*	1		
South	0.2431	0.4796*	0.6692*	1	
Islands	0.7689*	0.6686*	0.3307*	0.0580*	1

*Notes:** significant at 10%

Table A5: U/V Ratios Variance Decomposition

	(1)	(2)
VARIABLES	UVcollege	UVhighSchool
Constant	0.383**	0.936**
	(0.044)	(0.099)
sigma_u	0.3986	1.000
sigma_e	0.419	0.9486
rho	0.4750	0.5265
Observations	90	90
# macroarea	5	5

Notes: The U/V ratios are measured for the 5 Italian macroareas in the years 1993 to 2010. $sigma_u$ is the between groups (macroareas) variation; $sigma_e$ is the within groups variation. *Rho* is the fraction of total variation due to between groups variation. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table A6: Alternative Measures of Overeducation for College Graduates

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
UV_univ	0.0211**	0.00563	0.00897	0.0303***	0.0114	0.00309
	(0.0102)	(0.0137)	(0.0100)	(0.00966)	(0.0110)	(0.00834)
UV_univ*YSG	-0.00291**	-0.00295**	-0.00341***	-0.00208*	-0.00255**	-0.00149***
	(0.00125)	(0.00119)	(0.000556)	(0.00112)	(0.00107)	(0.000456)
UV_contemporaneous	yes	yes	no	yes	yes	no
macroarea FE	no	yes	no	no	yes	no
macroarea*year FE	no	no	yes	no	no	yes
Observations	71,498	71,498	71,498	71,498	71,498	71,498
R-squared	0.047	0.048	0.049	0.020	0.021	0.021

Notes: Columns (1) to (3): outcome variable is 1 if at least 60% of workers employed in same profession has less than college degree. Columns (4) to (6): outcome variable is 1 if the individual is employed in low-skilled professions (groups 4 to 9 of 1-digit ISTAT professions). All U/V ratios are in logs. Each estimation includes controls for: gender, log of age, marital status, and potential years of experience. Robust standard errors are reported in parenthesis (* significant at 10%, ** significant at 5%, *** significant at 1%). Standard errors are clustered at graduation cohort-macroarea level.

Table A7: Alternative Measures of Overeducation for High School Graduates

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
UV_HighSchool	0.0259***	-0.000253	-0.00667	0.000868	0.00433	-0.00863	0.0152***	0.00533	0.000210	0.00319	0.000564	-0.00415
	(0.00619)	(0.00790)	(0.00824)	(0.0145)	(0.0111)	(0.0107)	(0.00379)	(0.00384)	(0.00377)	(0.00713)	(0.00554)	(0.00467)
UV_HighSchool*YSG	-0.000288	-0.00105	-0.000175	0.000614	-0.00268	-0.000597	-0.000608*	-0.000913***	-0.000313	0.00146*	-8.26e-05	0.000579
	(0.000639)	(0.000667)	(0.000470)	(0.00181)	(0.00171)	(0.00143)	(0.000318)	(0.000319)	(0.000234)	(0.000834)	(0.000864)	(0.000604)
UV_univ				0.0366**	0.00624	0.00790				0.0113*	0.00533	0.00619
				(0.0139)	(0.0132)	(0.0130)				(0.00655)	(0.00649)	(0.00625)
UV_univ*YSG				-0.00134	0.00108	0.000608				-0.00180**	-0.000744	-0.00109
				(0.00174)	(0.00174)	(0.00167)				(0.000774)	(0.000821)	(0.000731)
UV_contemporaneous	yes	yes	no	yes	yes	no	yes	yes	no	yes	yes	no
macroarea FE	no	yes	no	no	yes	no	no	yes	no	no	yes	no
macroarea*year FE	no	no	yes	no	no	yes	no	no	yes	no	no	yes
Observations	86,047	86,047	86,047	86,047	86,047	86,047	86,047	86,047	86,047	86,047	86,047	86,047
R-squared	0.047	0.049	0.049	0.048	0.049	0.049	0.007	0.008	0.008	0.007	0.008	0.008

Notes: Columns (1) to (6): outcome variable is 1 if at least 60% of workers employed in same profession has less than high school degree. Columns (7) to (12): outcome variable is 1 if the individual is employed in elementary occupations (group 8 of 1-digit ISTAT professions). All U/V ratios are in logs. Each estimation includes controls for: gender, log of age, marital status, and potential years of experience. Robust standard errors are reported in parenthesis (* significant at 10%, ** significant at 5%, *** significant at 1%). Standard errors are clustered at graduation cohort-macroarea level.

Table A8: High School Graduates Full Sample

Panel A. Dependent Variable: Employment Status

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
UV-HighSchool	-0.0898***	-0.0480***	-0.0308***	-0.0572**	-0.0407***	-0.0139
	(0.00741)	(0.00712)	(0.00531)	(0.0228)	(0.0146)	(0.0122)
UV_HighSchool*YSG	0.00556***	0.00673***	0.00453***	0.00204	0.00571***	0.00157
	(0.000626)	(0.000543)	(0.000343)	(0.00236)	(0.00204)	(0.00182)
UV_univ				-0.0288	-0.0179	-0.0283*
				(0.0191)	(0.0153)	(0.0161)
UV_univ*YSG				0.00394*	0.00177	0.00353*
				(0.00207)	(0.00182)	(0.00192)
UV_contemporaneous	yes	yes	no	yes	yes	no
macroarea FE	no	yes	no	no	yes	no
macroarea*year FE	no	no	yes	no	no	yes
Observations	134,430	134,430	134,430	134,430	134,430	134,430
R-squared	0.117	0.122	0.122	0.118	0.122	0.122

Panel B. Dependent Variable: Monthly Salary

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
UV-HighSchool	-0.0707***	-0.00970	-0.0178	0.0433	-0.00296	-0.0223
	(0.0111)	(0.0124)	(0.0107)	(0.0292)	(0.0185)	(0.0142)
UV_HighSchool*YSG	0.000618	0.00145**	0.00208***	-0.00995***	0.000698	0.00323**
	(0.000991)	(0.000596)	(0.000469)	(0.00286)	(0.00204)	(0.00130)
UV_univ				-0.0191	0.00542	0.00287
				(0.0227)	(0.0160)	(0.0159)
UV_univ*YSG				0.00375*	-0.00146	-0.00130
				(0.00213)	(0.00128)	(0.00125)
UV_contemporaneous	yes	yes	no	yes	yes	no
macroarea FE	no	yes	no	no	yes	no
macroarea*year FE	no	no	yes	no	no	yes
Observations	30,742	30,742	30,742	30,742	30,742	30,742
R-squared	0.191	0.203	0.203	0.195	0.203	0.203

Panel C. Dependent Variable: Overeducation

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
UV_HighSchool	0.0451***	0.0186**	0.0102	0.0257	0.0284**	0.00993
	(0.00669)	(0.00917)	(0.00937)	(0.0186)	(0.0135)	(0.0122)
UV_HighSchool*YSG	-0.00142**	-0.00210***	-0.00101**	0.000338	-0.00341*	-0.000609
	(0.000614)	(0.000585)	(0.000423)	(0.00203)	(0.00174)	(0.00141)
UV_univ				0.0354**	-0.00795	-0.00200
				(0.0172)	(0.0148)	(0.0146)
UV_univ*YSG				-0.00206	0.000433	-0.000539
				(0.00193)	(0.00169)	(0.00162)
UV_contemporaneous	yes	yes	no	yes	yes	no
macroarea FE	no	yes	no	no	yes	no
macroarea*year FE	no	no	yes	no	no	yes
Observations	117,011	117,011	117,011	117,011	117,011	117,011
R-squared	0.049	0.052	0.052	0.050	0.052	0.052

Notes: The sample is including both the sub-sample of individuals who report the exact year of graduation and the sub-sample of individuals who report age at graduation from high school. In the presence of measurement error due to possible recall error (which causes abnormal spike at 19 in the age of graduation) the estimates are biased towards zero and are less precise, relative to the same analysis performed on the first sub-sample only (table 3.6). All U/V ratios are in logs. Each estimation includes controls for: gender, log of age, marital status, and potential years of experience. Robust standard errors are reported in parentheses (* significant at 10%, ** significant at 5%, *** significant at 1%). Standard errors are clustered at graduation cohort-macroarea level.

Bibliography

Akerlof, G. A. and Kranton, R. E. (2002). Identity and Schooling: Some Lessons for the Economics of Education. *Journal of Economic Literature*, 40(4):1167–1201.

Altonji, J. G., Blom, E., and Meghir, C. (2012). Heterogeneity in Human Capital Investments: High School Curriculum, College Major, and Careers. *NBER Working Papers*, page 58.

Altonji, J. G., Kahn, L. B., and Speer, J. D. (2016). Cashier or Consultant? Entry Labor Market Conditions, Field of Study, and Career Success. *Journal of Labor Economics*, 34(S1):361–401.

Andreoni, J. and Vesterlund, L. (2001). Which is the fair sex? gender differences in altruism. *The Quarterly Journal of Economics*, 116(1):293–312.

Anelli, M. and Peri, G. (2015a). Gender gap in italy: The role of college majors. In *Unexplored Dimensions of Discrimination*, pages 79–109. Oxford University Press.

Anelli, M. and Peri, G. (2015b). More unexplored dimensions of gender gap and college choice: Attitudes, choice of partner and peer/teacher effects in school. In *Unexplored Dimensions of Discrimination*, pages 110–128. Oxford University Press.

Azmat, G. and Petrongolo, B. (2014). Gender and the labor market: What have we learned from field and lab experiments? *Labour Economics*, 30:32–40.

Barnichon, R. and Zylberberg, Y. (Forthcoming 2018). Under-employment and the trickle-down of unemployment. *American Economic Journal: Macroeconomics*.

Becker, G. (1994). *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education (3rd Edition)*. National Bureau of Economic Research, Inc.

Bertrand, M., Goldin, C., and Katz, L. F. (2010). Dynamics of the gender gap for young professionals in the financial and corporate sectors. *American Economic Journal: Applied Economics*, 2(3):228–55.

Black, S. and Devereux, P. (2011). Recent developments in intergenerational mobility. volume 4B, chapter 16, pages 1487–1541. Elsevier, 1 edition.

Booth, A. and Nolen, P. (2009). Gender differences in risk behaviour: Does nurture matter? IZA Discussion Papers 4026, Institute for the Study of Labor (IZA).

Braga, M. and Checchi, D. (2008). Closing the Gender Gap? Life Competences and Social Environment. *Rivista di Politica Economica*, 98(5):155–198.

Card, D. and Payne, A. A. (2017). High school choices and the gender gap in STEM. Technical report.

Cheng, A., Kopotic, K., and Zamarro, G. (2017). Can parents' growth mindset and role modelling address stem gender gaps? Technical report, EDRE Working Paper No. 2017-07.

Clark, A. E., Georgellis, Y., and Sanfey, P. (2001). Scarring: The Psychological Impact of Past Unemployment. *Economica*, 68(270):221–241.

Dellas, H. and Koubi, V. (2003). Business cycles and schooling. *European Journal of Political Economy*, 19(4):843–859.

Destefanis, S. and Fonseca, R. (2007). Matching efficiency and labour market reform in italy: A macroeconometric assessment. *LABOUR*, 21(1):57–84.

Di Pietro, G. (2006). Regional labour market conditions and university dropout rates: Evidence from italy. *Regional Studies*, 40(6):617–630.

Eckel, C. and Grossman, P. (1998). Are women less selfish than men? evidence from dictator experiments. *Economic Journal*, 108(448):726–35.

Farré, L. and Vella, F. (2012). The intergenerational transmission of gender role attitudes and its implications for female labour force participation. *Economica*, 80(318):219–247.

Flabbi, L. (2012). Gender Differences in Education, Career Choices and Labor Market Outcomes on a Sample of OECD Countries. *World Development Report 2012*.

Fortin, N. M. (2008). The Gender Wage Gap among Young Adults in the United States: The Importance of Money versus People. *Journal of Human Resources*, 43(4):884–918.

Frühwirth-Schnatter, S., Pamminger, C., Weber, A., and Winter-Ebmer, R. (2010). Labor market entry and earnings dynamics: Bayesian inference using mixtures-of-experts markov chain clustering. NRN working papers 2010-14, The Austrian Center for Labor Economics and the Analysis of the Welfare State, Johannes Kepler University Linz, Austria.

Gelbach, J. B. (2016). When do covariates matter? and which ones, and how much? *Journal of Labor Economics*, 34(2):509–543.

Genda, Y., Kondo, A., and Ohta, S. (2010). Long-term effects of a recession at labor market entry in japan and the united states. *Journal of Human Resources*, 45(1):157–196.

Gibbons, R. and Waldman, M. (2006). Enriching a theory of wage and promotion dynamics inside firms. *Journal of Labor Economics*, 24(1):59–107.

Gneezy, U., Niederle, M., and Rustichini, A. (2003). Performance in competitive environments: Gender differences. *The Quarterly Journal of Economics*, 118(3):1049–1074.

González de San Román, A. and de la Rica, S. (2012). Gender Gaps in PISA Test Scores: The Impact of Social Norms and the Mother's Transmission of Role Attitudes. IZA Discussion Papers 6338, Institute for the Study of Labor (IZA).

Gottfried, M. A. and Bozick, R. (2016). Supporting the STEM pipeline: Linking applied STEM course-taking in high school to declaring a STEM major in college. *Education Finance and Policy*, 11(2):177–202.

Guiso, L., Monte, F., Sapienza, P., and Zingales, L. (2008). DIVERSITY: Culture, gender, and math. *Science*, 320(5880):1164–1165.

Hagedorn, M. and Manovskii, I. (2013). Job selection and wages over the business cycle. *American Economic Review*, 103(2):771–803.

Isfol (2005). *La domanda di lavoro qualificato: le inserzioni a modulo nel 2003*. I libri del Fondo sociale europeo. ISFOL.

Legewie, J. and DiPrete, T. (2014). Pathways to Science and Engineering Bachelor's Degrees for Men and Women. *Sociological Science*, 1(February):41–48.

Leuven, E. and Oosterbeek, H. (2011). Overeducation and mismatch in the labor market. IZA Discussion Papers 5523, Institute for the Study of Labor (IZA).

Liu, K., Salvanes, K. G., and Sørensen, E. Ø. (2016). Good skills in bad times : Cyclical skill mismatch and the long-term effects of graduating in a recession. *European Economic Review*, 84:3–17.

Lockwood, B. (1991). Information externalities in the labour market and the duration of unemployment. *Review of Economic Studies*, 58(4):733–753.

Niederle, M., Segal, C., and Vesterlund, L. (2013). How Costly Is Diversity? Affirmative Action in Light of Gender Differences in Competitiveness. *Management Science*, 59(1):1–16.

Nollenberger, N., Rodríguez-Planas, N., and Sevilla, A. (2016). The math gender gap: The role of culture. *American Economic Review*, 106(5):257–61.

OECD (2015a). Gender wage gap. Online; accessed February 2018.

OECD (2015b). Graduates by field. Online; accessed February 2018.

Oreopoulos, P., von Wachter, T., and Heisz, A. (2012). The short- and long-term career effects of graduating in a recession. *American Economic Journal: Applied Economics*, 4(1):1–29.

Osikominu, A., Grossmann, V., and Osterfeld, M. (2014). Are sociocultural factors important for studying a science university major? Annual conference 2014 (hamburg): Evidence-based economic policy, Verein für Socialpolitik / German Economic Association.

Oyer, P. (2006). Initial labor market conditions and long-term outcomes for economists. *Journal of Economic Perspectives*, 20(3):143–160.

Paglin, M. and Rufolo, A. M. (2016). NORC at the University of Chicago Heterogeneous Human Capital , Occupational Choice , and Male-Female Earnings Differences Published by : The University of Chicago Press on behalf of the Society of Labor Economists and the NORC at the University of Chicag. 8(1):123–144.

Petrongolo, B. and Pissarides, C. A. (2001). Looking into the black box: A survey of the matching function. *Journal of Economic Literature*, 39(2):390–431.

Philippis, M. D. (2017). STEM graduates and secondary school curriculum: does early exposure to science matter? Temi di discussione (Economic working papers) 1107, Bank of Italy, Economic Research and International Relations Area.

Piazzalunga, D. (2018). The gender wage gap among college graduates in italy. *Italian Economic Journal*.

Sahin, A., Song, J., Topa, G., and Violante, G. L. (2014). Mismatch unemployment. *American Economic Review*, 104(11):3529–64.

Wiswall, M. and Zafar, B. (2015). Determinants of college major choice: Identification using an information experiment. *Review of Economic Studies*, 82(2):791–824.

Xie, Y. and Schauman, K. A. (2003). *Women in science: Career processes and outcomes*. Cambridge: Harvard University Press.

Zafar, B. (2013). College major choice and the gender gap. *Journal of Human Resources*, 48(3):545–595.