Available online at www.sciencedirect.com

ScienceDirect

Journal of Policy Modeling 45 (2023) 494–521

www.elsevier.com/locate/jpm

Early Influences and the choice of college major: Can policies reduce the gender gap in scientific curricula (STEM)?[☆]

Silvia Granato [a,b,1,2]

[a] *European Commission, Joint Research Centre (JRC)*
[b] *University of Warwick, UK*

Received 12 October 2022; Received in revised form 10 March 2023; Accepted 12 April 2023
Available online 26 April 2023

Abstract

I use rich administrative and survey data to analyse the determinants of the gender gap in STEM graduation rates among Italian college graduates. Results from both Gelbach and Oaxaca decompositions show that half of the gender gap in STEM graduation is attributed to the gender difference in maths and science content of the respective high school curricula. This finding indicates that in Italy the gender gap in STEM graduation has its roots in a gendered choice originating many years before and that effective interventions aimed at increasing girls' interests in science and technology should be implemented at an early stage.

[☆] I would like to thank: Erich Battistin and Barbara Petrongolo for their precious guidance; the seminar participants in the Queen Mary University of London reading group and the University of Modena seminar series, and participants in the AIEL 2019 conference and the Workshop in Gender and Economics at the University of Luxembourg, for relevant comments and suggestions. I also thank the AlmaLaurea staff for making the data available and for the support received during my stay at their institution. The scientific output expressed does not imply a policy position of the European Commission. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use that might be made of this publication

E-mail address: silvia.granato@ec.europa.eu.

[1] *Address:* Via E. Fermi, 2749, 21027 Ispra (VA). The outputs of the empirical analyses presented in this paper have been produced while employed at the University of Warwick. Subsequent revisions of the text together with minor revisions of the empirical analyses have been carried on while working at the JRC of the European Commission.

[2] A previous version is contained in the author's PhD thesis: "Gender Inequalities and Scarring Effects in School to Work Transitions" (2018), available at http://qmro.qmul.ac.uk/xmlui/handle/123456789/43950.

https://doi.org/10.1016/j.jpolmod.2023.04.006

JEL Classification: I20; J16

Keywords: STEM; Education; Gender

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. In 2020, across all OECD countries, women were greatly over-represented among bachelor's degree graduates in the education field (in which the females share was 77%), but they represented only 26% of engineering graduates and 22% of graduates in the ICT field.[3]

Women's under-representation in science, technology, engineering and mathematics (STEM) studies is both perceived as a barrier to a sufficient supply of STEM skills[4] –which are considered crucial to boosting innovation and growth (see, for example, the evidence for US from Winters, 2014; Peri et al., 2016)– and an important contributor to the gender wage gap –because STEM degrees typically lead to higher-paying jobs and the gender difference in the choice of college major translates into gender gaps in earnings later in life (see the evidence from Machin & Puhani, 2003; Flabbi, 2012; Anelli & Peri, 2015; Card & Payne, 2021).

Why do women choose STEM degrees to a lower extent relative to men? To answer this question, a growing economic literature is investigating the determinants of the STEM gender gap. Females and males may select themselves in different university majors because of different preparation and achievement at pre-collegiate levels of education: if females acquire lower science and maths ability and knowledge prior to choosing their major this potentially translates in a lower probability of choosing a STEM degree. On the other hand, it may be the case that skills determine major choice differently for women and men, so that women are less likely to choose STEM even conditional on the same maths and science preparation upon college entry. Both different pre-college preparation and skills as well as different choices conditional on these skills can be the result of environmental influences, which can originate from a close environment, such as the family, or from the broader social setting in which students grow up and make their choices. Furthermore, if higher education supply varies across geographical location and females are less mobile than men, for example because of stronger family ties, then STEM higher education supply might play a role in explaining women's lower probability of choosing a STEM degree. In this paper I analyse the determinants of the gender gap in STEM graduation rates for Italian college- leaving cohorts from 2010 to 2015, with an emphasis on school, family and cultural influences, as well as geographic proximity in the supply of STEM degrees. I use data from a uniquely rich and largely unexplored source (AlmaLaurea) that combines administrative and survey information on the population of Italian

---

[3] Latest figure available from OECD (*Graduates by field*), available at https://stats.oecd.org.

[4] In 2020, the share of new entrants in tertiary education across OECD countries was only 27%. Latest figure available from OECD (*Distribution of new entrants by field of education*)

graduates. I am able to characterise the students' pre-college education in its most relevant aspects, including the curriculum of the high school attended and the student's grade at the high school final exam. 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. 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.

The unadjusted gender gap in STEM graduation rates for 2010–2015 cohorts is approximately 22 percentage points. I implement a Gelbach decomposition of this difference and show that its most important determinant, 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. Difference in high school performance and in family and social environments play a negligible role, so that, when this large set of characteristics is controlled for, half of the gap remains unexplained. In other words, even among students with identical measured pre-college factors, a 12 percentage points gap remains. The results from an Oaxaca-Blinder decomposition show that, even conditional on having attended one of the maths- and science-intensive high school tracks, girls have a much lower probability than boys of choosing a STEM degree. The results also suggest that family and social background features – over and above the influence they can already have on previous choices – affect female and male college choices differently, each accounting for another approximately 13% and 15%, respectively, of the STEM gender gap.

On the one hand, the results indicate that, in Italy, half of the STEM graduation gender gap may be tackled by addressing the gender difference in the choice of high school track. This could be achieved through interventions aimed at promoting females' participation in STEM fields already in middle school. On the other hand, it emerges that another relevant determinant of the STEM gender gap in higher education is parental influence, even conditional on other factors. Thus, measures targeting parents with the aim of promoting gender equality in the household and increasing their awareness of the influence they have on kids' choices might be effective in encouraging young females' participation in STEM careers.

I contribute to the literature on the gender gap in STEM in at least three ways. First, I add to the discussion on how much the gender gap in the choice of a STEM major is explained by different preparation at the end of secondary school relative to different choices of females and males with the same level of preparation at college entry. Previous evidence from Ireland (Delaney & Devereux, 2019) and Canada (Card & Payne, 2021) shows that, among college applicants, most of the gender gap in STEM enrolment results from differential subject choices among boys and girls in secondary school. The settings of both studies are characterised by early tracking, with secondary school students choosing which courses to take towards their final year and enrolment in higher education being conditional on grades in the chosen subjects. More recently, Speer (2021) shows that in the US –where there is no high school STEM track and the first pivotal stage where choices become more binding is the initial major choice– while differences in high school courses and grades among boys and girls enrolling in college are substantial, a largest portion of the gender gap in STEM majoring is explained by different choices conditional on high school preparation. I provide novel evidence from a different institutional framework, where subjects choice happens even earlier, when boys and girls at 14 years old decide to enrol into high school tracks that are more intensive in maths and science or in humanities, but high school track does not condition access to university in any field, which is mostly free. I show that in a context with early tracking into different secondary school types but free university admission, graduation in STEM programmes is still closely linked to high school preparation.

Second, I investigate the importance of early parental and societal influences in explaining the differential choice of a STEM major of boys and girls. While a vast body of literature demonstrates positive correlations between parents and children in terms of economic, educational, social, and behavioural outcomes, only very recently few studies have focused on the intergenerational transmission of STEM (Cheng et al., 2017; Bordón et al., 2020; Chise et al., 2021; Anaya et al., 2022). Similarly, several studies indicate a direct relationship between attitudes towards women and the maths gender gap in a given society (see Guiso et al., 2008; Nollenberger et al., 2016), but the evidence on the impact of individuals' sociocultural background on their choice of college major is scarce (see Osikominu et al., 2020). Thanks to the richness of the data used, I am able to measure relevant characteristics of the students' closer external environment, i.e. the family, and of the broader social settings in which they live, i.e. the civic community, and I estimate their direct effects on college major choice net of their influence on previous educational choices.

Finally, I provide novel evidence on the role of geographical proximity to the supply of university courses in influencing differently college major choice of females and males. Previous literature has documented lower geographical mobility of females relative to males in the labour market, mainly driven by stronger family ties (Shauman & Xie, 1996; Jürges, 2006; Farré & Ortega, 2021). Still, the role played by differential geographical mobility of females and males in the choice of college major remains largely unexplored. To the best of my knowledge, this is the first paper investigating the role of differential mobility across genders in influencing the choice of college major.

The remainder of the paper is organised as follows. Section 2 identifies the factors and mechanisms potentially shaping the gender gap in major choices, motivating the empirical analysis. Section 3 describes the institutional setting and the data used to measure the different relevant factors identified. Section 4 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 5 offers concluding remarks.

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 two sets of explanations: human capital factors – i.e., a student's preparation and achievement at pre-collegiate levels of education– and parental and societal influence – which can in turn affect both high school choices and individuals' preferences for higher education. Finally, I discuss the potential role of the local supply of higher education programmes.

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. 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 & Bozick, 2016; Tchuente, 2016; Philippis, 2021). Thus, if boys and girls have different preparation in terms of knowledge of science and maths acquired at high school, this may be reflected in different choices of enrolling in a STEM university course. The existing literature has investigated whether boys and girls make systematically different choices prior to college entry. Studies analysing high school students course choices in the US (Shauman & Xie, 2003; Speer, 2021) and Canada (Card & Payne, 2021) find that, on average, girls and boys take the same amount of maths and science high school courses, but girls are more likely to choose chemistry and biology and less likely to choose physics and calculus relative to their male counterparts. Despite this finding, interestingly, both Speer (2021) and Card and Payne (2021) show that among college applicants/enrolled students there is a significant gap in STEM 'readiness', defined as having certain requisites that make them ready to enrol in a STEM college major,[5] which both attribute to a differential selection into college, namely that non-STEM ready males attend college at much lower rates than non-STEM ready females.

Secondary education may also impact major choices via specific (observable or unobservable) high school characteristics, over and above their general track. Previous literature has demonstrated that pupils' outcomes are driven not only by their characteristics but also by school-level variables (see, for example Ferraro & Põder, 2018). For example, Legewie and DiPrete (2014) find that, all else being equal, gender segregation in extra-curricular activities has a discernible impact on the gender gap in the choice of a STEM major in the US. This evidence may be consistent with the self-selection of girls into high schools with certain characteristics predictive of STEM choice, or with a differential gender impact of such characteristics.

Conditional on high school choice, performance and final grades may play a role in STEM university enrolment. On the one hand, since STEM degrees are typically considered the most demanding ones (see, for example, Flabbi, 2012),[6] gender differences in high school overall performance may translate in a gender gap in the STEM choice at university. On the other hand, given the gender gap in maths ability largely documented by the literature (for a recent review, see Bertocchi & Bozzano, 2020), subject specific grades might be even more important in explaining the STEM enrolment and graduation gender gap. For example, Speer (2017) shows that, in the US, pre-college test score gaps in science and mechanical subjects explain more than twice as much of the gender gap in STEM majoring as the maths and verbal

---

[5] Speer (2021) defines STEM readiness based on ASVAB (Armed Services Vocational Aptitude Battery) test scores in science and maths knowledge tests (scoring at least one standard deviation above the age-adjusted mean) and high school course choice (having taken biology, chemistry, physics, calculus). Card and Payne (2021) measure STEM readiness as having completed at least three STEM related classes in the last year of high school.

[6] In a sample of higher education graduates from 14 OECD countries, Flabbi (2012) finds that science fields attract 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 the percentage of men and women who regard study programmes in the scientific field as very demanding is more than 20%, as opposed to only approximately 10% when they are asked about humanities programmes of study.

tests scores. Instead, Delaney and Devereux (2019) show that, among Irish students applying for college, male comparative advantage in maths in high school explains only a small part of the gender gap in the choice of a STEM field at university.

Even females and males with identical preparation at college entry might make different choices of major. In fact, findings suggest that skills are statistically weaker determinants of college major choice for women than men. In the US, Jiang (2021) finds that the marginal effect of both general academic skills and STEM-specific skills on the likelihood of graduating with a STEM degree is statistically larger for men than for their female counterparts. Similarly, Bordón et al. (2020) show that Chilean male students tend to apply to a higher extent to the most selective majors when they have good academic achievement, compared to female students with similar academic achievement.

2.2. *Family and social background*

Different pre-college preparation and skills as well as different choices conditional on these skills can both be the result of environmental influences. 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 the school, or from broader social settings in which individuals live, such as the civic community.

Concerning family influence, on the one hand, parents' educational achievement is important to the extent that it proxies for parents' abilities and skills, which are strong predictors of the abilities and skills of their children.[7] On the other hand, the family environment may be relevant for the transmission not only of skills but also of gender norms; in fact, several studies document a positive correlation between the gender role attitudes of parents and children.[8] Furthermore, the parental influence might be more or less strong depending on the gender of parents and children: other studies have demonstrated the existence of a gendered pattern in the influence of parents' education or occupation on children's educational and occupational choices (Pascual, 2009; Cheng et al., 2017; Bordón et al., 2020; Chise et al., 2021; Anaya et al., 2022).[9]

---

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

[8] 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 mothers' attitudes, which in turn influence female labour market decisions.

[9] Cheng et al. (2017) provide interesting evidence on maternal role modelling for daughters' choices: they find that having the mother employed in a STEM occupation increases the probability of the daughter working in hard sciences. In Chile, Bordón et al. (2020) find that male students are more strongly influenced by the field of occupation of the father relative to the one of the mother (in all fields), while for females students the father's influence is stronger in STEM fields and the mother's influence is stronger in non-STEM fields. Similarly, Chise et al. (2021) document that STEM-educated fathers are more important for sons, whereas STEM-educated mothers are more important for daughters; moreover, fathers with non-STEM degree have a negative influence on the probability that their sons complete a STEM degree, but not their daughters, while mothers with a non-STEM degree have a positive influence on their daughters' decision to pursue a STEM degree, but not of their sons. More recently, Anaya et al. (2022) study the

In addition, the civic community in which individuals grow up can be important for the transmission of gender norms, which may influence educational choices and achievement. Previous studies have provided evidence on the hypothesis that males' higher maths test scores reflect gender inequalities in a given culture.[10] Recently, Osikominu et al. (2020) have investigated the role of students' sociocultural background for choosing a STEM major at university in a sample of Swiss university graduates, for which they find that the gender gap in STEM is lower in less conservative municipalities, because male students from more conservative backgrounds are more likely to study STEM.

2.3. Higher education supply

Students' enrolment decision is potentially also a function of the supply of higher education that a student faces upon high school exit. This supply may vary across geographical location, if higher education institutions are not evenly geographically distributed. Previous literature has emphasized the relevance of the regional dimension in determining students' outcomes (see, for example, Agasisti & Cordero-Ferrera, 2013). A group of studies has analysed the determinants of students' mobility by means of estimation of gravity models, where students' flows to a given university are functions of university characteristics potentially impacting student mobility – measuring university *attractiveness* – and attitudes towards mobility – measured by the student-university distance (Sá et al., 2004; Dotti et al., 2014; Cattaneo et al., 2017). Females might be less likely than males to leave the family and move – because of different preferences or social attitudes towards females' choices. For example, recently Farré and Ortega (2021) have shown that Spanish female college graduates have stronger family ties than males, which restrict their geographical mobility and have a negative effect on their educational aspirations. This would imply that, given the same distance from a STEM course, females are less likely to enrol in such a course.

3. Institutional background and data

In the Italian education system, the first stage 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,[11] business and

---

(*footnote continued*)
role of maths achievement, self-perceived maths ability and parental occupation in explaining the probability of majoring in science in college. They find that all three factors are significant predictors of the probability of STEM majoring and that the parental influence (namely, the influence of having at least a parent working in a STEM related occupation) is concentrated among females.

[10] 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 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.

[11] This includes technical institutes offering specialization in: IT, electronics and electrical engineering, industrial chemistry, building engineering, nautics and areonautics.

accounting, administration, tourism, social services. In order to graduate from upper secondary school and receive a certificate that gives access to higher education, students have to pass a final exam that tests their knowledge of all the subjects studied. Students graduating from any high school have access to higher education with free choice of the major.

The Italian higher education system is organized in two cycles: a 3-year bachelor's degree, followed by a 2-year master's degree. For some specific fields, degree courses are organised in six- or five-year single cycle degree programmes (most of the health degree courses together with architecture and primary education sciences). Access to these programmes is regulated nationally and based on an entry test. For other degree courses, enrolment is mostly free, except for the fact that single higher education institutions can autonomously regulate access to their degree courses if they expect the demand to exceed their capacity.

3.1. STEM in the Italian context

In Italy, a list of the STEM university courses is provided by the Ministry of Education, University and Research (MIUR).[12] These are the courses that correspond to groups 5, 6 and 7 of the FOET (Fields of Education and Training) 2013 classification: natural sciences, mathematics and statistics, information and communication technologies, and engineering, manufacturing and construction. The STEM definition includes a fairly heterogeneous group of fields of study, namely: life sciences, physical sciences (including physics, chemistry, and earth sciences), maths & stats, computing, engineering, manufacturing, architecture and town planning, and building and civil engineering.

Using administrative data from MIUR on students' enrolment in Italian universities and on the very detailed content of each of the approximately 2,500 unique undergraduate or single-cycle courses offered by Italian higher education institutions, I obtain the enrolment gender gap and the 'maths content' of each STEM and non-STEM field of study. The maths content is measured with the proportion of university 'credits' (measured according to the European Credits Transfer and Accumulation System) that students must obtain in maths-intensive subjects out of all the credits they need in order to graduate from a specific degree course. Fig. 1 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 service fields). The correlation between these two measures is −60%. Even at the level of more than 2,500 unique university courses, the correlation is almost −50%. Interestingly, the evidence suggests that, in Italy, the 'STEM gender gap' issue appears to be mainly a 'maths gender gap' issue.[13]

---

[12] Since 2020, the Ministry of University and Research (MUR) is separated from the Ministry of Education, with the latter focusing on school administration only.

[13] Similar evidence has been provided for the US. Kahn and Ginther (2017) document that women's under-representation is limited to the maths-intensive STEM fields, for which they introduce the acronym 'GEMP' (geosciences, engineering, economics, math/computer science, and physical science), distinguishing them from the other STEM fields indicated with the acronym 'LPS' (life sciences, psychology, and social sciences excluding economics). Cimpian et al. (2020) report that the male-to-female ratio among US college majors is largely more unbalanced in physics, engineering, and computer science (indicated with the acronym 'PECS') than in the remaining STEM fields.

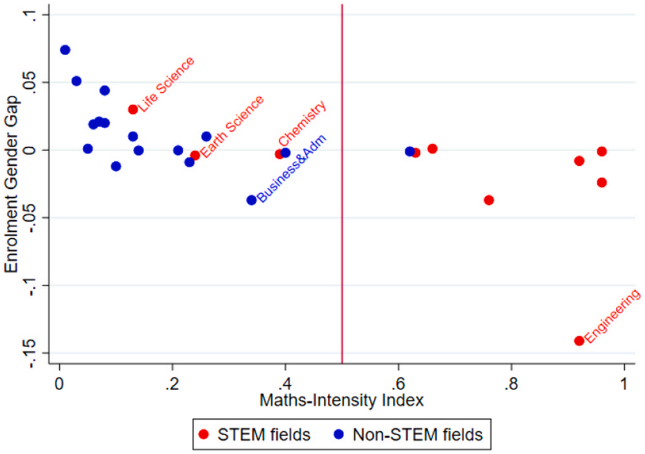


Fig. 1. Enrolment gender gap and maths intensity by fields of education. *Notes:* Each observation is a field of study. The average maths intensity across all courses in a given field is represented on the x-axis, while the y-axis shows the female-male difference in the probability of enrolling in each field.

3.2. Data

3.2.1. Student data

To analyse students' choices of major, I use data from the *AlmaLaurea Graduates' Profile*, a dataset combining administrative and survey information on the population of college graduates from a large sample of Italian universities, which is made available by the inter-university consortium AlmaLaurea. I focus on students from undergraduate and single-cycle courses graduating from 2010 to 2015 from one of the 56 universities taking part in the survey for the whole period considered (approximately 63% of all Italian higher education institutions). This sample covers approximately 65% of the population of the 2010–2015 cohorts of the Italian college graduates, and it is representative of the reference population across gender and field of study (see Table A1 in the Appendix.)[14]

Administrative variables provided by single institutions include rich information on students' university career –including students' degree course, which allows identifying whether the student graduated from a STEM field – and pre-college experience. The latter include high

---

[14] Not all students enrolled in universities will obtain a degree, and in this sense, the AlmaLaurea database represents only a selected sample of higher education students. In particular, if the drop-out or field-switching rate is differential between male and female students, this might result in an over- or under-estimation of the gender gap in the initial choice of enrolling in a STEM major (for example, Astorne-Figari and Speer, 2018 find a large gender gap in STEM persistence in the US, driven by a much larger probability of women starting STEM to switch major relative to their male peers). In Fig. A1 in the Appendix, I compare aggregate graduation rates by gender and year of enrolment obtained from the AlmaLaurea data with data on enrolment rates in STEM fields by gender and year of enrolment obtained from MIUR (enrolment data are available from MIUR for the years since 2003, only aggregated at the university, field of study and province of residence level). The figure documents the absence of differential drop-out/switching rates among females and males, 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. Nonetheless, 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 from 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.

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–, high school final grade and the names of the specific high schools attended by each student –which allows controlling for the role of other high school characteristics over and above their general track. The dataset includes information from a survey administered to all students upon graduation, with response rates above 90% for all cohorts. To characterise a student's family background, I draw on answers to questions about the level and field of education of both parents and their last occupation to proxy for socio-economic status.

3.2.2. *Local variables from other data sources*

I draw on the information on students' municipality of residence upon university enrolment to characterise a student's sociocultural background at the time of major choice. 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. Moreover, following Braga and Checchi (2008), I use as proxies for women's sexual emancipation the municipality-specific fertility rate[15] 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. On average, only 10% of the municipalities are governed by a female mayor, and these are concentrated in the northern part of the country. The average share of female councillors in local governments is only 20% and the percentage is higher in northern municipalities. The average fertility rate is approximately 39, with no clear and sharp geographical pattern. Finally, most marriages in Italy are celebrated with religious rituals (68%, on average) and the rate is higher in southern Italy (see Fig. A2 in the Appendix).

3.2.3. *Supply of STEM education*

In order to obtain a measure of the availability of STEM courses offered to prospective tertiary education students, I use administrative data made available by MIUR to measure the different factors characterising higher education supply in Italy, and I summarise them in a single *supply index*. In particular, for each STEM and non-STEM degree course available, I extract information on: the geographical location (Italian municipality) in which it is offered, the number of students enrolled yearly, and the percentage of need-based scholarships awarded to eligible students at the university offering the course. This information is used to construct a *supply index* for each Italian municipality, obtained by summing the number of courses – both overall and of STEM fields only– offered in all Italian municipalities, weighted by: (i) the geographical proximity to the municipality where the course is offered –as measured by the physical distance between the student's municipality of origin and each degree course municipality; (ii) the size of the university offering the course –proxied by the number of students enrolled yearly in each course; and (iii) the percentage of scholarships awarded to eligible

---

[15] The fertility rate is calculated as the number of live births in a municipality divided by the number of women between ages 15 and 49, multiplied by 1,000.

students at each university –capturing the heterogeneity in pecuniary costs faced by students enrolling at different higher education institutions.[16]

Fig. 2 is a plot of the resulting index for the year 2010 for the supply of all courses (panel a) and STEM only courses (panel b) 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 to students coming from southern regions, and this variation may account for differences in STEM graduation rates between students from different regions of the country.

3.3. Final sample and summary statistics

The number of college graduates from 2010 to 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. I focus on 3-year undergraduate or 5-year single cycle students, further restricting the sample to students who were born in Italy and residing in Italy at graduation and who enrolled between the ages of 18 and 21 in the years from 2003 to 2011, i.e. the years for which I have data on the variables measured at municipal level. The final sample consists of 504,673 observations.

Table 1 lists summary statistics of the main variables, presented separately for male and female students in the sample. Females constitute 61% 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 the maths intensity of the course chosen, I find a gender gap that is similar in magnitude: the percentage of maths-intensive subjects in degree courses chosen by females is, on average, 21 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.[17] Moreover, females are much less likely than men to have graduated from a technical STEM high school. Thus, the majority of men are tracked early on into classes with higher exposure to science and maths, and viceversa for girls. On the other hand, females outperform males: on average, they obtain a higher final high school grade. When looking at family characteristics, i.e. parents' level of education and type of job, females appear to have parents who are slightly less educated and have lower-level jobs

---

[16] The geographical proximity weight has value 1 if the course is offered in the same municipality or in municipalities within 1 km linear distance, and it is the inverse of the linear distance for courses offered in other municipalities. The weight for size is increasing in the following four size categories: very large (more than 40,000 students enrolled), large (between 20,000 and 40,000 students enrolled), medium (between 10,000 and 20,000 students enrolled), and small (less than 10,000 students enrolled). The weight for scholarships availability is the percentage of scholarships awarded to eligible students. The main results of the paper are robust to alternative approaches to the construction of the index (i.e., using different ways of calculating the weights for the three components of the index) as well as to different approaches to the inclusion of the supply measures in the estimation (i.e., including the three components of the index separately in the regression specification). These results are not reported for sake of brevity and are available from the author upon request.

[17] The Scientific and Technical STEM categories are indicators for having attended respectively a 'scientific' academic high school offering students a maths- and science-intensive curriculum or a 'technical' vocational high school offering specialisation in technological subjects such as IT, electronics or chemistry. The non-STEM category includes humanities-intensive academic high schools including 'classics', 'languages' and 'artistic' tracks and non-STEM technical vocational high schools with specialisation in business and administration or tourism.

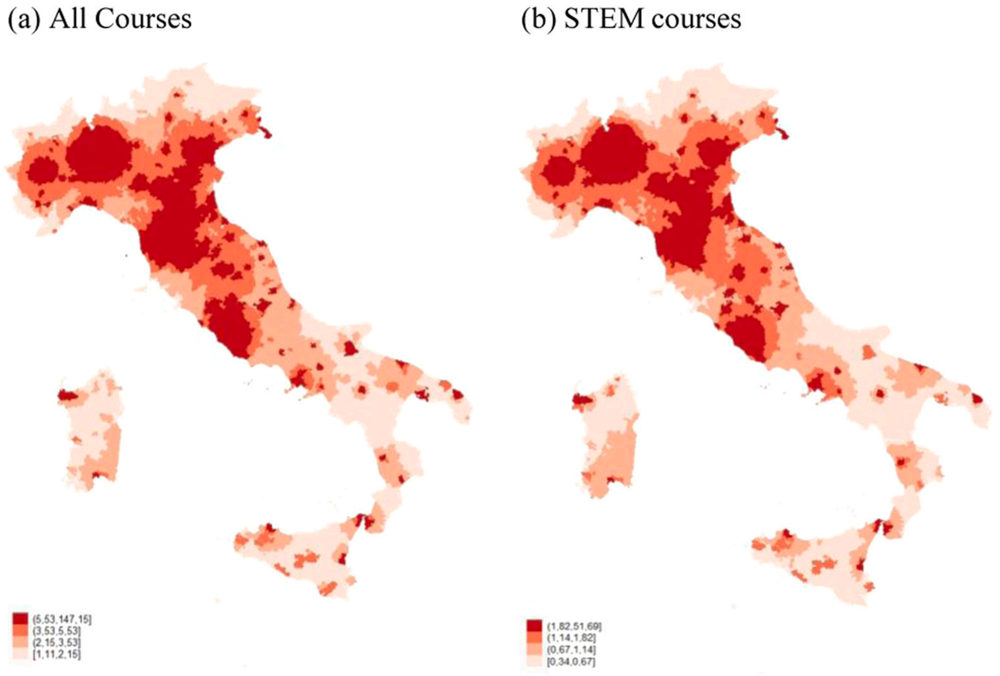


Fig. 2. Supply index. *Notes:* The two panels plot the index of supply in 2010, obtained for each municipality by summing the number of all/STEM-only courses offered in all other municipalities, weighted by the linear distance, the size of the university offering the course and the percentage of scholarships awarded by each university.

with respect to their male counterpart. Finally, there is no difference between females and males in the variables measured in the municipality of residence in the year of enrolment at university, indicating that, as expected, male and female graduates are equally distributed in the national territory.

4. Empirical analysis

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

$$y_{im\tau t} = \beta_1 F_i + X_i\beta_2 + Z_{m\tau}\beta_3 + \eta_t + \gamma_m + \delta_\tau + u_{im\tau t}$$

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 binary variable taking value 1 if student i is female; 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. Year and municipality of enrolment fixed effects, as well as dummies for the survey year (corresponding to the students' year of graduation) are included in the model.

Given the estimate of the gender gap in the outcome, $\hat{\beta}_1$, in order to identify and discuss the contributions from each of the four groups of variables – pre-college education, family characteristics, social background and the supply of STEM degree courses – I adopt the conditional decomposition suggested by Gelbach (2016), which allows obtaining, for each covariate, a

Table 1
Summary Statistics.

Variables	Males		Females	
	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:				
Science	0.54	0.50	0.38	0.49
Technical STEM	0.17	0.38	0.02	0.13
Non-STEM	0.29	0.45	0.60	0.49
Final grade	80.68	12.44	83.69	12.03
Family Characteristics:				
Father education:				
Less than HS	0.29	0.46	0.37	0.48
HS	0.46	0.50	0.44	0.50
College non-STEM	0.17	0.37	0.13	0.34
College STEM	0.08	0.27	0.06	0.23
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 STEM	0.04	0.20	0.03	0.17
Father last occupation:				
Blue collar (or never worked)	0.16	0.37	0.21	0.41
Self-employed/small businessman	0.14	0.35	0.17	0.38
White collar	0.41	0.49	0.37	0.48
Liberal professions/white collar director/entrepreneur	0.29	0.45	0.24	0.43
Mother last occupation:				
Housewife	0.23	0.42	0.26	0.44
Blue collar	0.10	0.30	0.13	0.33
Self-employed/small businessman	0.08	0.27	0.09	0.29
White collar	0.50	0.50	0.45	0.50
Liberal professions/white collar director/entrepreneur	0.09	0.29	0.08	0.26
Municipality Characteristics:				
Fertility Rate	39.22	7.19	39.07	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	7.83	16.02	6.91	15.00
Supply of university courses	24.55	49.76	21.75	46.68
Observations	192,010		312,663	

Notes: Sample includes 3-year undergraduate or 5-year single-cycle students who enrolled between 2003 and 2011 and graduated between 2010 and 2015.

parameter measuring its contribution in explaining the gender gap in the outcome, which is the female-male gap in the value of the variable scaled by its STEM graduation equation impact.[18]

---

[18] Given the equation of the base model: $y_{im\tau t} = \alpha_0 + \alpha_1 F_i$ 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 (1) given by the omitted variable bias formula. The difference $(\hat{\alpha}_1 - \hat{\beta}_1)$ 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 indicator.

Table 2
Gelbach Coefficient Decomposition.

Outcome	Graduated in STEM	
	(1)	(2)
Estimated STEM gender gap	-0.217***	-0.217***
	(0.003)	(0.003)
HS curriculum:		
3 categories	-0.098***	
	(0.002)	
High school fixed effects		-0.104***
		(0.002)
HS performance	0.012***	0.012***
	(0.000)	(0.000)
Family characteristics	-0.003***	-0.002***
	(0.000)	(0.000)
Municipal variables	0.000***	0.000***
	(0.000)	(0.000)
Supply	0.001**	0.002**
	(0.001)	(0.001)
Cohort FE	0.000***	0.000***
	(0.000)	(0.000)
Municipality FE	-0.002***	-0.003**
	(0.001)	(0.001)
Full regression coefficient	-0.127***	-0.123***
	(0.002)	(0.002)
Observations	504,673	504,673
R squared	0.189	0.239

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

Notes: Decompositions of the gender gap in STEM graduation rate 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 binary variable equal to 1 if the individual graduated from a STEM field. In column (1) high school curriculum is measured in 3 categories: non-STEM, scientific and technical STEM; in column (2) the high school curriculum variables group includes more than 11,000 identifiers for secondary institution and track attended. High school performance is measured in 3 categories for the intervals 60–85, 85–95, and 95–100. All other groups of variables are defined as in Table 1. Each regression includes survey year, year of graduation and municipality of residence fixed effects. Standard errors are clustered at municipality of residence level.

Table 2 reports the results from this decomposition of the coefficient measuring the gender gap in STEM graduation. Column (1) displays results from the estimation of a model where the high school curriculum is included in 3 categories, namely non-STEM – which is the excluded category – scientific, and technical STEM. The high school track alone here explains approximately 45% of the gender gap in the outcome. The second most important contribution is given by high school performance: the positive female-male difference in high school final grade documented in the previous section contributes to reducing the gender gap in STEM college graduation by approximately 5.4%. All the remaining variables together account for less than 3% of the STEM gender gap.

In column (2) 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 the 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. Including the full set of school-track dummies leaves the results almost unchanged; thus, very little is due to differences in the 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, science and technology content of students' high school curriculum. Among college graduates, females and males have different preparation at the end of high school because of different educational choices already made at the age of 14, with girls less likely to enrol in high school tracks that are more intensive in maths, science and technology. This difference in high school choice shapes students' future education careers, despite being in a relatively open institutional setting, where high school track does not necessarily determine specific occupational opportunities.

Better high school performance is positively associated with graduating from a STEM degree, and girls on average complete high school with a higher final grade. This evidence implies that if girls were under-performing relative to boys in high school, the gender gap in major choices would be even greater.

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. Even among females and males with identical high school preparation, family and cultural background and geographical proximity to STEM degree courses, there remains a gender gap in the probability of graduating in STEM of approximately 12 percentage points.

4.1. The role of different returns to individual characteristics

The analysis based on the estimation of model (1) assumes that the coefficients on the covariates are not different 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 preferred specification, i.e. the model that includes the high school track in 3 categories. 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 are presented in Table 3; all the predictors included in the regressions are summarised in five groups, as done above.

Column (2) reports the terms indicating the contribution of each group of variables to the part of the gap explained by differences in observed characteristics between females and males. The results confirm that the high school curriculum contributes the most to the portion of the gap due to differences in characteristics, followed by the high school performance.[19] All the other terms are of negligible magnitude.

Column (3) reports the coefficients indicating the different contribution of each group of variables to the portion of the gender gap in the outcome accounted for by gender differences in returns to observed characteristics. A sizable proportion (approximately one quarter) is attributed to different returns to the high school track: even conditional on the high school track attended, females have lower probabilities of choosing a STEM degree.

---

[19] It is worth noticing that the coefficients of the 'explained' portion of the gap 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.

Table 3
Oaxaca Decomposition.

	Differential	Explained	Unexplained
	(1)	(2)	(3)
Females	0.173***		
	(0.003)		
Males	0.39***		
	(0.003)		
Difference	-0.217***		
	(0.003)		
High School Track		-0.109***	-0.034***
		(0.002)	(0.001)
High School Performance		0.018***	-0.055***
		(0.000)	(0.001)
Family Characteristics		-0.001**	0.028***
		(0.000)	(0.004)
Municipal Variables		0.000	-0.032**
		(0.000)	(0.015)
Supply Indexes		0.000	0.005*
		(0.000)	(0.003)
Constant			-0.037**
			(0.016)
Total		-0.092***	-0.125***
		(0.002)	(0.002)

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

Notes: Oaxaca decomposition of the gender gap in STEM graduation rate of university courses. The sample consists of college graduates enrolled in 2003–2010 and graduated in 2010–2015 (504,673 observations). The dependent variable is a binary variable equal to 1 if the individual graduated from a STEM field. Each regression includes survey year, year of graduation and municipality of residence fixed effects. The other variables are defined as in Table 2.

Interestingly, different returns to high school performance play an even bigger role: girls are less sensitive than boys to their level of overall skills, measured by high school final grade, when deciding to enrol in STEM fields, and this accounts for almost half of the ‘unexplained’ portion of the STEM graduation gender gap. As shown in Table A2 – which reports the full regression results – having a top final mark (95–100) from high school relative to a lower final mark (60–86) increases girls’ probability of graduating in STEM by only 9 percentage points compared to 22 percentage points for males. This result is consistent with the recent evidence from other countries documenting that girls are less sensitive to their level of skills when enrolling in STEM at university.

Furthermore, it emerges from column (3) that differences in returns to family characteristics and socio-cultural background matter in explaining the gender gap in STEM graduation, each accounting for another approximately 13% and 15%, respectively, of the overall STEM gender gap. As it can be seen from Table A2 (columns 2 and 3), consistently with what found in other studies for Italy and other countries, the relationship between parents education and students’ STEM graduation varies according to the gender of both the child and the parent and the field of study. In particular, having either parent with a STEM college degree is associated with a higher probability of graduating from a STEM degree, with the association being stronger for father-son (vs father-daughter) and mother-daughter (vs mother-son). Having a father with tertiary non-STEM education is negatively correlated with the outcome and more strongly for male students, while the opposite is true for mothers, i.e. having a mother with a non-STEM degree is positively correlated with the outcome and more strongly for female students. Furthermore, on top of parents' level and type of education, the variables measuring their broad occupation group have mostly a non-significant correlation with the outcome, except for the case of the parent being in the liberal profession/white collar director/entrepreneur group –which has a negative correlation with the STEM graduation probability, significant only for males. Liberal professions are mostly non-STEM, such as doctors or lawyers, and the results seem to suggest that the son, not the daughter, is more likely to follow the profession of the parent.[20]

When looking at variables measuring students' socio-cultural background, we observe (Table A2) that the only relevant covariate is the share of religious marriages, which is negatively correlated with the probability of choosing a STEM degree only for females. This result might suggest 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.

Finally, the results indicate no significant differential response of female and males to geographical proximity to supply of STEM higher education is found. This evidence indicates no role for differential mobility of females and males – suggested by previous literature – in explaining the gender gap in STEM graduation, net of the other measured influences.

4.2. *Heterogeneity of results across socio-economic status*

One may wonder whether the results on the role of the different measured factors in influencing students' choice of major are heterogeneous across students with different socio-economic background.

When performing the Gelbach and Oaxaca decompositions, respectively, of the gender gap in STEM graduation rates on three sub-samples defined according to the socio-economic status of the student's family,[21] it emerges that, while the gender gap in major choices declines with socio-economic status,[22] the role of the different groups of variables in explaining the gap does not appear to differ significantly across the three sub-samples (Table A3 and Table A4 in the Appendix).

Overall, we see that 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.

---

[20] These dynamics for Italy are explored more deeply in Chise et al. (2021) and Aina and Nicoletti (2018).

[21] The variable measuring socio-economic status is constructed based on the answers of students to questions regarding their parents' last occupation. 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; and high – parents who are directors or owners of businesses with at least 15 workers or who are self-employed in liberal professions. The social group of the family refers to the highest between the two parents.

[22] This result is driven by both females' probability of graduating from STEM programmes increasing with social status and the opposite for males. This evidence may be consistent with the hypothesis that in families where the parents are employed in liberal professions, which are typically non-STEM, (the group with the higher socio-economic background according to the definition adopted in this section) the sons tend to follow the profession of the parents and are less likely to choose STEM majors.

5. 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. The under-representation of women in STEM fields contributes both to the scarcity of scientists across developed countries and to the gender wage gap in the labour market. Understanding the mechanisms underlying the educational segregation of women is policy relevant as it may help better shaping the policies aimed at encouraging females' participation in STEM careers.

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

The first key finding is 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 academic high school and the technical vocational high school with a focus on industrial construction and preparation for surveyors, which are the fields that ensure the highest returns to STEM enrolment in college. This finding is relevant from a policy perspective, because it indicates that in Italy the STEM gender gap in higher education has its roots in a gendered choice that has already taken place many years before. It emerges that for Italian students the high school choice, despite not being binding for university and occupational opportunities, is a pivotal stage that impacts their future educational paths. Thus, 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.

The second key finding is that, even conditional on the high school track choice, a relevant role is played by the different influences of the family and social backgrounds on the decisions of females and males of graduating from a STEM field. This finding suggests that the role of environmental factors – such as the family and the broader social environment – in the different educational choices of females and males is even greater than its direct net impact on college major choice that can be estimated through this study. Thus, from a policy perspective, early interventions targeting female students alone might not be sufficient to fully address the STEM gender gap. Policymakers might consider combining them with measures targeting parents of lower or upper secondary students, aimed at both promoting more equal gender norms in the household and increasing parents' awareness of the influence they have on children's choices.

Appendix

See Appendix Fig. A1, Fig. A2 and Tables A1 – A4.

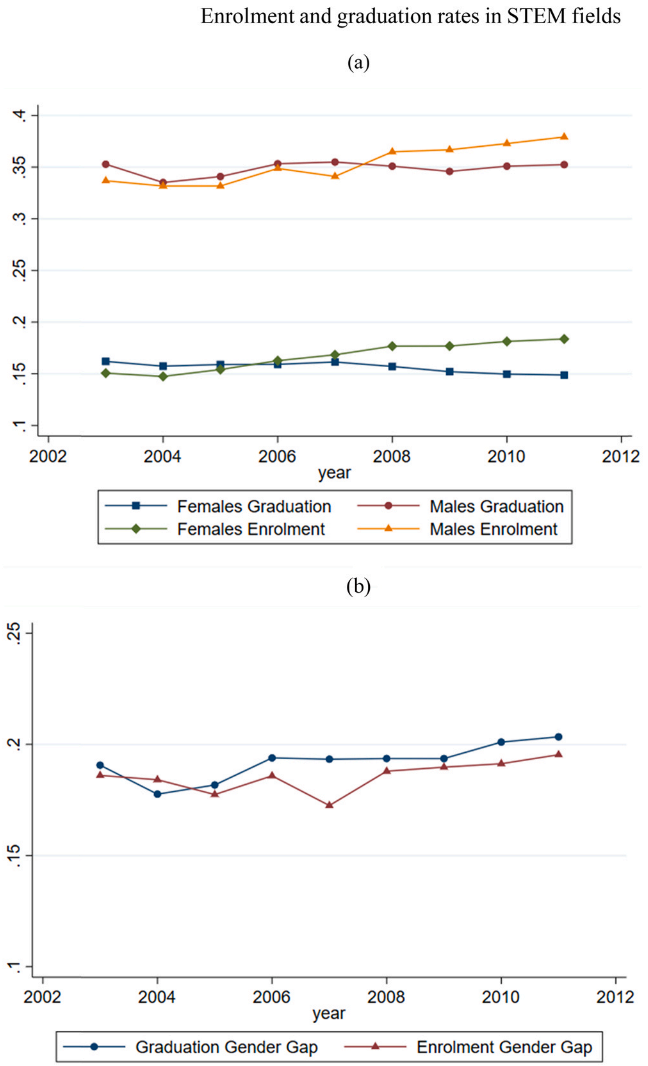
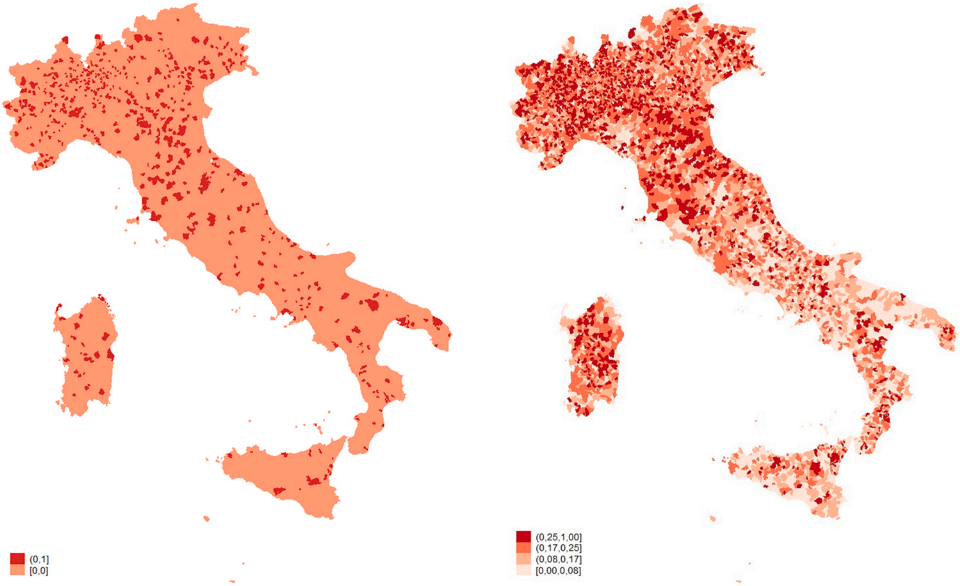


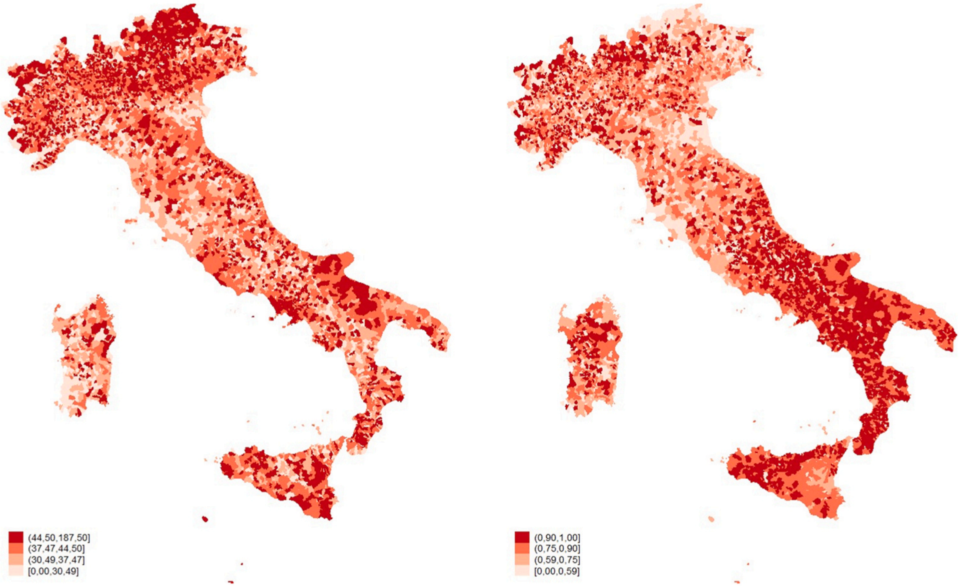
Fig. A1. Enrolment and graduation rates in STEM fields. *Notes:* Enrolment rates (number of students enrolled in STEM fields as a percentage of the total number of students enrolled) are obtained from MIUR data for students enrolled in an undergraduate or single-cycle degree programme between 2003 and 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 degree programme between 2010 and 2015 and who enrolled between 2003 and 2012, from universities taking part in the AlmaLaurea survey from 2010.

Municipal variables



(a) Female Mayors

(b) Share of Female Councillors



(c) Fertility Rate

(d) Share of Religious Marriages

Fig. A2. Municipal variables. *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*.

Table A1
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 A2
Full Regressions: STEM Graduation Rate.

	Pooled	Males	Females	Pooled	Males	Females
Variables:	(1)	(2)	(3)	(4)	(5)	(6)
female	-0.127***			-0.123***		
	(0.00226)			(0.00232)		
High School track (humanities excluded):						
Science	0.208***	0.265***	0.178***			
	(0.00211)	(0.00366)	(0.00225)			
Technical STEM0	0.429***	0.457***	0.406***			
	(0.00379)	(0.00472)	(0.00806)			
High school final grade:						
85–95	0.0931***	0.153***	0.0551***	0.0913***	0.151***	0.0549***
	(0.00181)	(0.00303)	(0.00193)	(0.00191)	(0.00316)	(0.00199)
95–100	0.141***	0.225***	0.0917***	0.140***	0.226***	0.0921***
	(0.00238)	(0.00375)	(0.00232)	(0.00263)	(0.00403)	(0.00252)
Father education (less than HS excluded):						
High school	0.0122***	0.0106***	0.0137***	0.00852***	0.00894***	0.00866***
	(0.00152)	(0.00301)	(0.00157)	(0.00150)	(0.00310)	(0.00158)
College non stem	-0.0264***	-0.0508***	-0.00492	-0.0314***	-0.0515***	-0.0125***
	(0.00316)	(0.00498)	(0.00310)	(0.00266)	(0.00462)	(0.00275)
College stem	0.124***	0.135***	0.113***	0.113***	0.126***	0.102***
	(0.00424)	(0.00603)	(0.00485)	(0.00363)	(0.00586)	(0.00447)
Mother education (less than HS excluded):						
High school	0.00432***	1.24e-05	0.00664***	0.00186	0.000698	0.00274
	(0.00150)	(0.00291)	(0.00172)	(0.00156)	(0.00300)	(0.00182)
College non stem	0.00784***	0.00461	0.0118***	0.00207	0.00261	0.00401
	(0.00272)	(0.00475)	(0.00269)	(0.00232)	(0.00437)	(0.00257)
College stem	0.0836***	0.0749***	0.0902***	0.0743***	0.0694***	0.0782***
	(0.00559)	(0.00684)	(0.00725)	(0.00434)	(0.00624)	(0.00610)

(*continued on next page*)

Table A2 (*continued*)

	Pooled	Males	Females	Pooled	Males	Females
Father last occupation (blue collar or never worked excluded):						
Self-employed/small businessman	0.00569***	0.00488	0.00623***	0.00405**	0.00428	0.00451**
	(0.00194)	(0.00404)	(0.00210)	(0.00186)	(0.00398)	(0.00211)
White collar	0.00354*	5.79e-05	0.00434**	0.00215	0.000300	0.00316
	(0.00181)	(0.00351)	(0.00193)	(0.00179)	(0.00365)	(0.00192)
Liberal professions/white collar director/entrepreneur	0.000894	-0.0116***	0.00918***	-0.00175	-0.0129***	0.00603**
	(0.00205)	(0.00381)	(0.00244)	(0.00201)	(0.00384)	(0.00250)
Mother last occupation (housewife excluded):						
Blue collar	-0.00891***	-0.0117***	-0.00704***	-0.00954***	-0.0142***	-0.00780***
	(0.00212)	(0.00427)	(0.00228)	(0.00208)	(0.00434)	(0.00230)
Self-employed/small businessman	0.00254	-0.00841*	0.00898***	0.000678	-0.00802*	0.00599**
	(0.00247)	(0.00436)	(0.00294)	(0.00241)	(0.00435)	(0.00290)
White collar	0.00144	-0.00884***	0.00751***	-0.00162	-0.0112***	0.00468**
	(0.00169)	(0.00320)	(0.00200)	(0.00164)	(0.00309)	(0.00201)
Liberal professions/white collar director/entrepreneur	-0.0106***	-0.0256***	-0.000443	-0.0133***	-0.0248***	-0.00565**
	(0.00256)	(0.00456)	(0.00298)	(0.00236)	(0.00454)	(0.00282)
Municipal Variables						
Female mayor	0.00294	0.00433	0.00164	0.00334	0.00212	0.00377
	(0.00509)	(0.0101)	(0.00418)	(0.00497)	(0.00907)	(0.00429)
Share female councillors	0.00994	0.00775	0.0159	0.00818	0.0129	0.0101
	(0.0172)	(0.0243)	(0.0182)	(0.0175)	(0.0251)	(0.0181)
Fertility rate	0.000153	0.000443*	3.64e-05	0.000143	0.000537*	4.11e-05
	(0.000143)	(0.000269)	(0.000158)	(0.000142)	(0.000278)	(0.000159)
Share of religious marriages	-0.00558	0.0176	-0.0159*	-0.00321	0.0209	-0.0147*
	(0.00737)	(0.0139)	(0.00819)	(0.00731)	(0.0141)	(0.00821)
Supply of STEM courses	-0.00381	-0.00541**	-0.00257	-0.00341	-0.00553**	-0.00245
	(0.00277)	(0.00237)	(0.00360)	(0.00307)	(0.00255)	(0.00360)
Supply of university courses	0.000815	0.00121*	0.000542	0.000588	0.00120*	0.000354
	(0.000821)	(0.000734)	(0.00113)	(0.000882)	(0.000727)	(0.00112)
School fixed effects	No	No	No	Yes	Yes	Yes
Constant	0.146***	0.0822***	0.0472***	0.279***	0.305***	0.137***

(*continued on next page*)

Table A2 (*continued*)

	Pooled	Males	Females	Pooled	Males	Females
	(0.00906)	(0.0158)	(0.00888)	(0.00919)	(0.0167)	(0.00884)
Observations	504,673	191,314	312,395	504,673	189,010	310,922
R-squared	0.189	0.197	0.118	0.239	0.262	0.176

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

Notes: The sample consists of college graduates who enrolled between 2003 and 2010 and graduated between 2010 and 2015. The dependent variable is a binary variable equal to 1 if the individual graduated from a STEM field. In columns (1)-(3) high school curriculum is measured in 3 categories: non-STEM, scientific and technical STEM; in columns (4)-(6) the high school curriculum variables group includes more than 11,000 identifiers for secondary institution and track attended. High school performance is measured in 3 categories for the intervals 60–85, 85–95, and 95–100. All other groups of variables are defined as in Table 1. Each regression includes survey year, year of graduation and municipality of residence fixed effects. Standard errors are clustered at municipality of residence level.

Table A3
Gelbach Decomposition by SES sub-samples.

Socio-economic status:	Low	Medium	High
	(1)	(2)	(3)
Estimated STEM gender gap	-0.281***	-0.23***	-0.166***
	(0.004)	(0.003)	(0.004)
HS curriculum (3 cateogories)	-0.133***	-0.103***	-0.072***
	(0.002)	(0.001)	(0.002)
HS performance	0.007***	0.012***	0.012***
	(0.001)	(0.000)	(0.000)
Parents	-0.001***	-0.003***	-0.003***
	(0.000)	(0.000)	(0.000)
Municipal variables	0.000***	0.000	0.000**
	(0.000)	(0.000)	(0.000)
Supply	0.001*	0.001**	0.001***
	(0.001)	(0.000)	(0.000)
Cohort fe	-0.002***	0.000***	0.002***
	(0.000)	(0.000)	(0.000)
Municipality FE	-0.002*	-0.002***	-0.002***
	(0.001)	(0.001)	(0.001)
Full regression coefficient	-0.151***	-0.135***	-0.103***
	(0.004)	(0.002)	(0.003)
Observations	64,532	292,462	145,327
R squared	0.286	0.201	0.182

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

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). All the groups of variables are defined as in Table 2, column (1).

Table A4
Oaxaca Decomposition by SES sub-samples.

Socio-Economic status:	Differential	Low Explained	Unexplained	Differential	Medium Explained	Unexplained	Differential	High Explained	Unexplained
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Females	0.136***			0.174***			0.191***		
	(0.002)			(0.002)			(0.004)		
Males	0.417***			0.405***			0.356***		
	(0.004)			(0.003)			(0.003)		
Difference	-0.281***			-0.23***			-0.166***		
	(0.004)			(0.003)			(0.003)		
Total		-0.132***	-0.149***		-0.098***	-0.132***		-0.062***	-0.103***
		(0.003)	(0.004)		(0.002)	(0.002)		(0.002)	(0.003)
HS curriculum		-0.142***	-0.038***		-0.115***	-0.039***		-0.078***	-0.027***
		(0.003)	(0.003)		(0.002)	(0.002)		(0.002)	(0.002)
HS performance		0.01***	-0.061***		0.018***	-0.059***		0.018***	-0.045***
		(0.001)	(0.004)		(0.001)	(0.002)		(0.001)	(0.002)
Parents education		-0.001**	-0.002		-0.001***	0.012***		-0.002***	0.029***
		(0.000)	(0.003)		(0.000)	(0.003)		(0.001)	(0.007)
Municipal variables		0.000	0.003		0.000	-0.047**		0.000	-0.054
		(0.000)	(0.046)		(0.000)	(0.021)		(0.000)	(0.033)
Supply		0.000	0.007		0.000	0.002		0.000	0.012**
		(0.000)	(0.008)		(0.000)	(0.003)		(0.000)	(0.005)
Constant			-0.058			-0.001			-0.018
			(0.048)			(0.021)			(0.034)
Observations	64,532	64,532	64,532	292,462	292,462	292,462	145,327	145,327	145,327

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

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 (low/medium/high).

References

Agasisti, T., & Cordero-Ferrera, J. M. (2013). Educational disparities across regions: A multilevel analysis for Italy and Spain. *Journal of Policy Modeling, 35*(6), 1079–1102.

Aina, C., & Nicoletti, C. (2018). The intergenerational transmission of liberal professions. *Labour Economics, 51*(December 2016), 108–120.

Akerlof, G. a, & Kranton, R. E. (2002). Some identity lessons of schooling: for the economics education. *Journal of Economic Literature, 40*(4), 1167–1201.

Anaya, L., Stafford, F., & Zamarro, G. (2022). Gender gaps in math performance, perceived mathematical ability and college stem education: the role of parental occupation. *Education Economics, 30*(2), 113–128.

Anelli, M., & Peri, G. (2015). Gender gap in Italy: The role of college majors. In T. Boeri, E. Patacchini, & G. Peri (Eds.). *Unexplored Dimensions of Discrimination*. Oxford University Press, Oxford.

Astorne-Figari, C., & Speer, J. D. (2018). Drop out, switch majors, or persist? The contrasting gender gaps. *Economics Letters, 164*, 82–85.

Bertocchi, G., & Bozzano, M. (2020). *Gender gaps in education. Handbook of Labor, Human Resources and Population Economics*. Springer International Publishing,1–31.

Black, S., & Devereux, P. (2011). Recent developments in intergenerational mobility. In O. Ashenfelter, & D. Card (Vol. Eds.), (1 edition..,). *Handbook of Labor Economics: volume 4B*, (pp. 1487–1541). Elsevier.

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

Braga, M., & Checchi, D. (2008). Closing the gender gap? Life competences and social environment. *Rivista di Politica Economica, 98*(5), 155–198.

Card, D., & Payne, A. A. (2021). High school choices and the gender gap in stem. *Economic Inquiry, 59*(1), 9–28.

Cattaneo, M., Malighetti, P., Meoli, M., & Paleari, S. (2017). University spatial competition for students: the Italian case. *Regional Studies, 51*(5), 750–764.

Cheng, A., Kopotic, K., & Zamarro, G. (2017). Can parentsà growth mindset and role modelling address STEM gender gaps? *SSRN Electronic Journal*.

Chise, D., Fort, M., & Monfardini, C. (2021). On the Intergenerational Transmission of STEM Education among Graduate Students. *B E Journal of Economic Analysis and Policy, 21*(1), 115–145.

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

Delaney, J. M., & Devereux, P. J. (2019). Understanding gender differences in STEM: Evidence from college applications. *Economics of Education Review, 72*(June), 219–238.

Dotti, N. F., Fratesi, U., Lenzi, C., & Percoco, M. (2014). Local labour market conditions and the spatial mobility of science and technology university students: evidence from Italy. *Review of Regional Research, 34*(2), 119–137.

Farré, L., & Ortega, F. (2021). Family ties, geographic mobility and the gender gap in academic aspirations. *SSRN Electronic Journal*(14561)).

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

Ferraro, S., & Põder, K. (2018). School-level policies and the efficiency and equity trade-off in education. *Journal of Policy Modeling, 40*(5), 1022–1037.

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

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

Gottfried, M. A., & 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. (2008). Differences in Test Scores Correlated with Indicators of Gender Equality. *Science, 320*(May), 1–2.

Jiang, X. (2021). Women in STEM: Ability, preference, and value. *Labour Economics, 70*(March), Article 101991.

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

Kahn, S., & Ginther, D. (2017). Women and science, technology, engineering, and mathematics (STEM). In S. L. Averett, L. M. Argys, & S. D. Hoffman (Eds.). *The Oxford Handbook of Women and the Economy* (pp. 766–798). Oxford: Oxford University Press.

Legewie, J., & DiPrete, T. A. (2014). Pathways to science and engineering bachelor's degrees for men and women. *Sociological Science, 1*(February), 41–48.

Machin, S., & Puhani, P. A. (2003). Subject of degree and the gender wage differential: Evidence from the UK and Germany. *Economics Letters, 79*(3), 393–400.

Nollenberger, N., Rodr´ıguez-Planas, N., & Sevilla, A. (2016). The math gender gap: The role of culture. *American Economic Review, 106*(5), 257–261.

Osikominu, A., Grossmann, V., & Osterfeld, M. (2020). Sociocultural background and choice of STEM majors at university. *Oxford Economic Papers, 72*(2), 347–369.

Pascual, M. (2009). Intergenerational income mobility: The transmission of socio-economic status in spain. *Journal of Policy Modeling, 31*(6), 835–846.

Peri, G., Shih, K., & Sparber, C. (2016). STEM workers, H-1B Visas, and productivity in US cities. *The Economics of International Migration, 49*(3), 277–307.

Philippis, M. D. (2021). STEM graduates and secondary school curriculum: Does early exposure to science matter? *Journal of Human Resources,* 1219–10624R1.

Sá, C., Florax, R. J., & Rietveld, P. (2004). Determinants of the regional demand for higher education in the Netherlands: A gravity model approach. *Regional Studies, 38*(4), 375–392.

Shauman, K., & Xie, Y. (1996). Geographic mobility of scientists: Sex differences and family constraints. *Demography, 33*(4), 455–468.

Shauman, K., & Xie, Y. (2003). *Women in Science: Career Processes and Outcomes.* Massachusetts, and London, England: Harvard University Press, Cambridge,.

Speer, J. D. (2017). The gender gap in college major: Revisiting the role of pre-college factors. *Labour Economics, 44*(December 2016), 69–88.

Speer, J. D. (2021). Bye Bye Ms. American Sci: Women and the Leaky Stem Pipeline. *SSRN Electronic Journal* (14676)).

Tchuente, G. (2016). High school human capital portfolio and college outcomes. *Journal of Human Capital, 10*(3), 267–302.

Winters, J. V. (2014). STEM graduates, human capital externalities, and wages in the U.S. *Regional Science and Urban Economics, 48*, 190–198.