Università degli Studi di Palermo

Dottorato in Scienze Economiche e Statistiche

Dipartimento di Scienze Economiche, Aziendali e Statistiche

SECS-S/05 - Statistica Sociale

Inequalities in student performances in the Italian universities

CANDIDATO

Andrea Priulla

COORDINATORE

Andrea Consiglio

TUTOR

Massimo Attanasio

CO-TUTOR

Martina Vittorietti

CICLO XXXV

ANNO CONSEGUIMENTO TITOLO 2023

Acknowledgements

This thesis is the result of a long journey in which many people have contributed to making it entertaining and challenging. First, I would like to thank my supervisor Massimo Attanasio for his trust in me from the beginning and for the continuous and helpful advices without which I could not have achieved this result. His passion for research and knowledge have been a source of inspiration for me. I would also like to thank my co-supervisor, Martina Vittorietti and Prof. Marco Enea, two of the smartest researchers I have had the good fortune to collaborate with over the recent years.

I thank my parents for their continuous support during these years and for helping me become who I am today. I thank Alessandro, Nicoletta, Totò, Furio, Andrea, and Daniele, a fantastic group who made this journey less long and stressful than it actually was. I especially thank Alessandro for convincing me to decide to enroll in statistics with him, and for driving me on his motorbike to university for eight years (many thanks also to his motorbike, Berty), and Nicoletta for waking me up (almost) every morning to study for the exams and for stealing all my snacks.

Finally, I would like to thank Stefania, the light at the end of this journey, for her patience over these years and for the patience she will have in the near future.

Contents

List of Tables **iv**

List of Figures **vii**

Introduction **1**

1 Does taking additional Maths classes in high school affect academic outcomes? **12**

1.1 Introduction 13

1.1.1 Theoretical Framework 15

1.1.2 The scientific "liceo" 16

1.2 Data description 18

1.3 Preliminary analysis 19

1.3.1 Track-level analysis 20

1.3.2 Student-level analysis 20

1.4 Methods 23

1.4.1 Multi-level propensity score matching 26

1.4.2 The discrete-time multi-state Markov model 28

1.5 Results 30

1.6 Conclusions 35

2 Analysis of university student mobility from the South to the Center-North of Italy **37**

2.1 Introduction 37

2.2 Student mobility in Italy at university enrollment 39

2.2.1 Inter-regional flows 39

2.2.2 The high school background 41

2.2.3 Central-northern universities: which are the most attractive? 44
2.3 Bachelor's degree achievement: who is faster? 46
2.3.1 The high school background 46
2.4 Mobility at master's degree enrollment 48
2.4.1 Inter-regional flows 48
2.4.2 The attractiveness of northern universities 50
2.5 Conclusions 56

3 An analysis of Italian university students' performance through segmented regression models: gender differences in STEM programs 58
3.1 Introduction 59
3.1.1 Theoretical framework 60
3.2 Data description 63
3.3 Preliminary analysis 64
3.4 Methods 66
3.4.1 Segmented regression models 68
3.4.2 Sequential hypothesis testing procedure for the choice of K_0 71
3.5 Results 74
3.6 Conclusions 78

4 Gender differences in university students' trajectories in Italy 83
4.1 Introduction 83
4.1.1 Theoretical framework 86
4.2 Data description 88
4.3 Preliminary analysis 89
4.3.1 The gender composition of the Italian universities 89
4.3.2 The transition from the first to the second year 91
4.3.3 University retention: who arrives at the end of the pipeline? 93
4.4 The discrete-time multi-state Markov model 99
4.5 Results 101
4.6 Conclusions 103

Conclusions 106

Limitations and Future Research 110

List of acronyms **112**

References **114**

CRediT Author Statement **128**

List of Tables

1.1 Number of hours per week devoted to scientific and non-scientific subjects in the two tracks of the scientific “liceo” since the 2010 reform. 18

1.2 Students’ characteristics of the two tracks of the scientific “liceo”. Cohort of the fifth-year high school students in 2018/19. 22

1.3 Estimated transition probabilities for both tracks of the scientific “liceo” before and after the MPSM procedure. CIs are in brackets. 32

2.1 Origin-destination (OD) matrix from the region of residence to the region of first university enrollment. University students enrolled in Italy in 2008/09. 42

2.2 Origin-destination (OD) matrix from the region of residence to the region of first university enrollment. University students enrolled in Italy in 2017/18. 43

2.3 AMR by macro-region of residence and type of high school attended. University students enrolled in 2008/09 and 2017/18. 44

2.4 BA graduation rates within four years by region of residence and mobility at enrollment. University students enrolled in Italy in 2008/09 and 2014/15. 47

2.5 BA graduation rates within four years by macro-region of residence and type of high school attended. University students enrolled in Italy in 2008/09 and 2014/15. 47

2.6 Mobility rates from the South to the Central-North regions in the school-BA and BA-MA transitions. Absolute values and percentages. University students enrolled in Italy in 2008/09 and 2014/15. 49

2.7 Origin-destination (OD) matrix from the region of bachelor’s degree within four years to the region of master’s enrollment within five years. University students enrolled in Italy in 2008/09. 51

2.8 Origin-destination (OD) matrix from the region of bachelor's degree within four years to the region of master's enrollment within five years. University students enrolled in Italy in 2014/15. 52

2.9 Origin (University of bachelor's graduation within four years) - destination (University of master's enrollment within five years) matrix. Absolute values and rate of stayers. University students enrolled in Italy in 2008/09. 53

2.10 Origin (University of bachelor's graduation within four years) - destination (University of master's enrollment within five years) matrix. Absolute values and rate of stayers. University students enrolled in Italy in 2014/15. 54

3.1 Female students enrolled for the first time in STEM programs, and female enrollment rate of the 2008 and 2014 cohorts. 65

3.2 Parameter estimates $\boldsymbol{\theta}$'s of the segmented regression model in Equation 3.5. Cohort of first-year students enrolled in 2014. Baselines are in brackets. . . 75

3.3 Parameter estimates of the $\boldsymbol{\psi}$'s and $\boldsymbol{\delta}$'s of the segmented regression model. Cohort of first-year students enrolled in 2014. 76

3.4 Number of selected changepoints in males and females credits in the course-specific segmented regression models using Davies' and Score tests. 76

3.5 Parameter estimates $\boldsymbol{\theta}$'s of the segmented regression models by university program. Baselines are in brackets. Cohort of first-year students enrolled in 2014 in biology, biotechnology, and chemistry. The asterisk indicates a corresponding p-value <0.05. Standard errors are in brackets. 79

3.6 Parameter estimates $\boldsymbol{\theta}$'s of the segmented regression models by university program. Baselines are in brackets. Cohort of first-year students enrolled in 2014 in computer science, engineering, and mathematics. The asterisk indicates a corresponding p-value <0.05. Standard errors are in brackets. . 80

3.7 Parameter estimates $\boldsymbol{\theta}$'s of the segmented regression models by university program. Baselines are in brackets. Cohort of first-year students in 2014 in natural sciences, physics, and statistics. The asterisk indicates a corresponding p-value <0.05. Standard errors are in brackets. 81

3.8 Parameter estimates of the $\boldsymbol{\psi}$'s and $\boldsymbol{\delta}$'s of the course-specific segmented regression models. Cohort of first-year students enrolled in 2014. Standard errors are in brackets. 82

4.1 Percentage of females enrolled in STEM and non-STEM programs at Italian universities. University students enrolled in Italy in 2008/09, 2014/15, and 2020/21. 90
4.2 Dropout rates and change of field of study rates in the transition from the first to the second year, by gender, the field of study, and the type of high school attended. University students enrolled in Italy in 2014/15. 92
4.3 BA degree rates within four years according to the field of study of enrollment and gender. The BA rate is calculated as the ratio between the total number of BA graduates within four years and the total number of BA enrollments. University students enrolled in Italy in 2008/09 and 2014/15. . 94
4.4 MA enrollment rates by gender and BA completion field of study. The MA rate is calculated as the ratio between the number of students enrolled at the MA level within five years and the total number of BA graduates within four years. University students enrolled in Italy in 2008/09 and 2014/15. . . 95
4.5 BA degree and MA enrollment rates by gender, the field of study, and macro-region. University students enrolled in Italy in 2014/15. 99

List of Figures

1.1 Students' ESCS distribution of the students enrolled at the fifth and last year of high school in Italy in 2018/19. 17

1.2 Percentages of traditional science or applied science tracks over the total number of tracks offered by high schools at the regional level in 2018/19. . . 21

1.3 School-university transition rates for the two science tracks at the regional level, calculated as the number of students enrolled at university in 2019/20 over the total number of fifth-year high school students enrolled in a specific region in 2018/19. 23

1.4 Percentage distribution of the students of the two tracks of the scientific "liceo" according to the university enrollment field of study. 24

1.5 Illustration of the two-step matching procedure. 25

1.6 States and transitions of the multi-state Markov model. 26

1.7 Difference in means between the two tracks of the scientific "liceo" of the track- and student-level covariates before (red) and after (black) the MPSM procedure. Positive values indicate an imbalance in favor of the applied science track, and negative values in favor of the traditional science track. . 31

1.8 Estimated probabilities for the *HS last-year* → *Not enrolled* transition by macroregion of the school, before and after the matching procedure. 33

1.9 Estimated probabilities for the *HS last-year* → *STEM* and *HS last-year* → *Non-STEM* transitions by macroregion of the school, before and after the matching procedure. 33

1.10 Estimated probabilities for the *STEM* → *0 credits* and *Non-STEM* → *0 credits* transitions by macroregion of the school, before and after the matching procedure. 34

1.11 Estimated probabilities for the *STEM* → *(30-60] credits* and *Non-STEM* → *(30-60] credits* transitions by macroregion of the school, before and after the matching procedure. 34

2.1 Inflows and outflows at enrollment in 3-year degree programs of Italian universities by region of residence. University students enrolled in Italy in 2008/09 and 2017/18. 40

2.2 Enrollment inflows from southern regions in the top eight attractive universities. University students enrolled in Italy in 2008/09 and 2017/18. 45

2.3 Outflows (negative values) and inflows (positive values) of the top ten southern universities in the transition to the master's level by macro-region of master's enrollment. Outflows are movers to other universities; inflows are movers from other universities. Absolute values. University students enrolled in Italy in 2008/09 and 2014/15. 55

2.4 Outflows (negative values) and inflows (positive values) of the top ten central and northern universities in the transition to the master's level. Outflows are movers to other universities; inflows are movers from other universities. Absolute values. University students enrolled in Italy in 2008/09 and 2014/15. 56

3.1 Median credits earned at the end of the first year (X-axis) by male and female students enrolled in STEM programs. Cohorts of first-year students enrolled from 2008 to 2014. 66

3.2 BA completion rates in STEM programs by gender and type of high school attended. Cohort of first-year students enrolled in 2014. 67

3.3 Simulated segmented relationship with a single changepoint. 69

3.4 Segmented relationships between the logit of the probability of success (Y-axis) and the credits earned at the end of the first year (X-axis) of the marginal model in Equation 3.4. Males are represented by the blue broken-line and females by the red one. Cohort of first-year students enrolled in 2014. 73

3.5 Segmented relationships between the logit of the probability of success (Y-axis) and the credits earned at the end of the first year (X-axis) of the program-specific segmented regression models in Equation 3.6. Males are represented by the blue broken-line and females by the red one. Cohort of first-year students enrolled in 2014. 78

4.1 Gender ratio (M/F) in non-STEM and STEM fields at BA enrollment and MA enrollment. University students enrolled in Italy from 2008/09 to 2014/15. 91

4.2 X-axis: dropout rates in the transition from the first to the second year; Y-axis: non-enrollment MA rates for BA in time graduates. The rates are calculated according to gender, the type of high school attended, and the field of study. The black straight lines represent the mean rates for each group. University students enrolled in Italy in 2014/15. 97

4.3 X-axis: dropout rates in the transition from the first to the second year; Y-axis: non-enrollment MA rates for BA graduates. By gender, STEM field of study, and type of high school attended. The black straight lines represent the mean rates for Xs and Ys in each group. University students enrolled in Italy in 2014/15. 98

4.4 X-axis: dropout rates in the transition from the first to the second year; Y-axis: non-enrollment MA rates for BA graduates. By gender, the non-STEM field of study, and the type of high school attended. The black straight lines represent the mean rates for Xs and Ys in each group. University students enrolled in Italy in 2014/15. 98

4.5 States and transitions of the multi-state Markov model. 100

4.6 Estimated probabilities for the enrolled → dropout transition. 103

4.7 Estimated probabilities for the enrolled → BA completion transition. 104

4.8 Estimated probabilities for the BA completion → MA enrollment transition. 104

Introduction

Since the Second World War, one of the most noticeable features of modern societies has been the expansion of educational opportunities and the increased demand for education at all levels. Three fundamental steps in the expansion of education occurred in the twentieth century: i) the introduction of nearly universal primary education, ii) the expansion of secondary education, and iii) the development of a system of mass higher education. Each step can be seen as a consequence of the previous one. The rise in secondary school enrollment resulted in a significant increase in the number of new entrants into tertiary education. The average share of 25-34 year-olds with a tertiary qualification increased from 27% in 2000 to 48% in 2021 across OECD countries. On average, tertiary education is now the most common attainment level among 25-34 year-olds and will soon be the most common among all working-age adults across the OECD countries (OECD, 2021).

Since educational expansion supports the development of democratic principles regarding status attainment, educational inequalities were expected to decrease (Hadjar and Becker, 2009; Kromydas, 2017). However, this growth has not been matched by the eradication of inequalities affecting the educational systems worldwide (Shavit and Blossfeld, 1993; Jackson, 2013; Bernardi and Ballarino, 2014). Inequalities in the chances of earning a university degree have increased over time among all social classes, which is entirely in contradiction with the expectations of theories of educational expansion due to democratizing access to higher education (Schizzerotto and Barone, 2006). Numerous studies have revealed that there is still a significant disparity in educational attainment among members of various groups. Most research focuses on inequalities related to students' social background, which is widely recognized as the primary factor contributing to educational inequalities. (Shavit and Blossfeld, 1993; Argentin and Triventi, 2011; Jackson, 2013; Contini et al., 2018). Further research has addressed the problem of inequalities related to socioeconomic status, gender, geographical location, and ethnic background (Clancy and Goastellec, 2007; Gibbons and Vignoles, 2012; Giudici et al., 2021).

In the sociological literature, the effects of the factors driving inequalities in educational choices and opportunities are usually distinguished into two distinct categories (Boudon, 1974): *primary* effects are all the effects related to educational outcomes, for example, in the relationship between social origin and high school performance; the decisions that children (and their families) make during their educational careers, such as the decision to leave, are *secondary* effects.

The Italian higher education system is characterized by high rates of early school dropout, low educational attainment, low school-to-university transition rates, and high student mobility rates. The most significant barriers to achieving an inclusive society where no one is left behind are territorial disparities, socioeconomic status, gender, and migrant status. This thesis looks into three types of inequalities in the Italian higher education system: socioeconomic inequalities in access and performance, gender differences in university choices and performance, and geographic differences in educational outcomes associated with student mobility. In this context of educational inequalities, this study assumes primary importance in several respects. First, it is helpful for universities to boost enrollments and identify students more at risk of early dropout by providing guidance courses to help them. Second, students' educational choices and outcomes are strictly linked to future job opportunities.

Socioeconomic inequalities

Educational outcomes are affected by an intricate net of factors, among which socioeconomic background plays a fundamental role (Checchi et al., 2003; Sirin, 2005; Ballarino and Panichella, 2016; Pensiero et al., 2019). Although the Universal Declaration of Human Rights Adopted and Proclaimed by the General Assembly of the United Nations on the Tenth Day of December 1948 states that everyone has the right to education, students from socioeconomically disadvantaged groups are still underrepresented (Mishra, 2020). Despite the detection of a decline of inequality in educational opportunities in several countries (Breen et al., 2009), the family background still represents an important predicting factor for individual attainments (Vergolini and Vlach, 2017). On average, students born into families with different cultural and socioeconomic resources have different levels of academic performance in primary school (Rosén et al., 2013). Both social background and early school performance may then work together in affecting the type of education received and students' performance in lower and upper secondary schools. In this respect, a school system is said to be horizontally stratified when students are allocated to different, non-overlapping curricula at each stage of their educational career. The common and most shared view in sociological literature establishes indeed that the more an educational system is stratified, the more social origin and parental background significantly influence the decision to pursue tertiary education.

In this regard, the high school track can be seen as a proxy for family wealth and among the main factors affecting students' educational careers. According to the effectively maintained inequality thesis (Lucas, 2001), socioeconomically advantaged families use their resources to secure some degree of advantage in the educational system for their children. Therefore, in secondary education, one can expect that children from the upper social strata will be more likely to attend the academic track, private schools, or high-ability classes that provide instruction conducive to university studies and, possibly, better labor market perspectives. For instance, in France, the proportion of students obtaining the upper secondary certificate (*baccalauréat*) has increased in the last years, with a strong stratification of the educational choices and opportunities. Lower-class students are over-represented in the technical and vocational types of *baccalauréat*, and upper-class students are over-represented in general *baccalauréat*, which leads to better opportunities for access to tertiary education (Duru-Bellat et al., 2008). A similar educational structure can be found in Germany, where the school system is highly stratified because of the early first transition point (Schneider, 2008; Neugebauer et al., 2013). In the German context, the educational decision made after elementary school is consequential for educational success. In this respect, the choice of the high school track is more restrictive than the Italian one concerning a potential engagement in tertiary education. Upper secondary school (*Gymnasium*) is the most common track and the main route leading to university enrollment. Although the *Gymnasium* is the main pathway for university enrolment, several additional but rarely used pathways could lead to eligibility for tertiary education (*Hauptschule*, *Realschule*, and *Gesamtschule*). However, completing additional programs is needed for university admission after graduating from these tracks.

According to Shavit and Blossfeld (1993), the relationship between social origin and educational attainment has remained stable in the last decades of the 20th century. More recent research highlights the tendency toward a weakening relationship between social class and educational attainment in many European countries (Jackson, 2013).

The horizontal stratification of the Italian school system implies an inevitable selection process of the students among the various curricula. Theoretically, this selection should be based on students' motivations. However, in Italy, more than in other countries (Checchi and Flabbi, 2007), this selection is based on the family's socioeconomic status, which is typically responsible for their children's educational choices (Contini and Triventi, 2016; Ballarino and Panichella, 2021). The increase in school attendance during the first decades of the twentieth century allowed elites to maintain their advantage by broadening the range of educational options. To meet the growing demand for education from the lower classes, the Gentile reform of the Italian educational system was proposed in 1923. That reform aimed to increase the horizontal stratification of secondary school to create a clearer distinction between elite and popular tracks (Schizzerotto and Barone, 2006). Basically, only students from classical high schools were eligible to enroll in any university faculty. Students who graduated from scientific high schools had access to technical and scientific faculties and medical fields. Finally, graduates of technical high schools and "istituti magistrali" were barred from enrolling in universities. In the second half of the 20th century, several reforms were proposed to reduce the hierarchy and stratification of the Italian educational system. There is no doubt that throughout the two decades between the early 1950s and the first half of the 1970s, the Italian educational system evolved into one that was less exclusive and more open than the one that resulted from the Gentile reform. For instance, 1969's reform extended eligibility to all individuals with five-year upper secondary education. This and other reforms enacted during the latter part of the 20th century did not, however, definitely modify its structure since transition rates are still very different across tracks.

Nowadays, the Italian high school system can still be described as a hierarchical tripartite system (Checchi et al., 2003): "licei", especially humanities and scientific ones, are the traditional schools preparing students for a potential university enrollment; technical schools, and all the technical tracks, are considered an intermediate choice between academic and vocational tracks; finally, vocational schools prepare students to directly enter the labor market and apply for various low-ability jobs (Contini and Scagni, 2013). Cappellari (2004) analyzed the transition from high school to university in Italy, confirming that students from traditional high schools, namely scientific and humanities "licei", are more likely to enroll at university and to perform better. Conversely, attending technical and vocational schools increases the employment probability.

Geographical inequalities

Since Italy's unification in 1861, politicians, economists, historians, and researchers have all been interested in the North-South disparity in the economic progress (Checchi and Peragine, 2005; Abramo et al., 2016). Many facets of our nation's structure reflect this regional division, and the higher education system is unquestionably one of them.
The territorial dualism is transferred to the universities in a vicious circle regarding social and economic development. The educational reforms implemented in Italy over the last 50 years have indirectly contributed to this divide. For example, Law 537 of 1993 established financial autonomy for universities to control costs in conjunction with increasing cuts in government spending. This reform has increased competition among Italian universities, contributing to growing disparities in services provided, research quality, and ability to attract students. This competitive environment among Italian universities has contributed significantly to "student mobility", i.e., the decision to leave the region of residence for study-related reasons, particularly during the transition from secondary school to university.

The moving can be vertical, towards economically more advanced and academically superior systems, or horizontal, between countries or institutions of more or less equal academic quality. Vertical mobility in Italy has a double meaning: vertical geographical mobility, observed from the southern regions to the northern ones, and vertical social mobility, in terms of improvement of employment and lifestyle (Vittorietti et al., 2022). Studies on Italian domestic students' mobility (Enea, 2016; Giambona et al., 2017; Attanasio et al., 2019) and graduates' mobility (Panichella, 2013; Iammarino and Marinelli, 2015) confirm that this is not only a matter of temporary mobility, but "Mezzogiorno" is experiencing a proper "brain drain" to the Center-North of Italy, resulting in further impoverishment of the human capital of the South. The mobility flows, driven by South-North disparities in work prospects and university quality, primarily target students with the highest academic results, which are associated with the socioeconomic background of their families of origin. In turn, their academic results, which are often less favorable than those reported in the Center-North, reflect students' lower level of preparation in southern universities, accentuated by outgoing mobility (Mariani and Torrini, 2022).

This has to be considered in relation to the dramatic dropout and low university enrolment rates that affect southern regions. In this regard, the work of Contini et al. (2018) has demonstrated that enrollment and retention are certainly lower in areas with high youth unemployment rates, such as the southern Italian regions, showing that when labor market prospects are poor, discouragement is prevalent.

Student mobility has been widely addressed in the literature in recent years. In Italy, student mobility is unidirectional, with southern students being attracted to the poles of attraction in the North-Center of our country (Columbu et al., 2021a). Genova et al. (2021) used network analysis to investigate student outflows from Sicily to other Italian regions, revealing the existence of some preferential paths reflecting the South-North orientation of student mobility also in the transition to the master's level. In detail, their findings show that Milano, Torino, and Bologna are the main attraction poles for Sicilian students both at university enrolment and at master's level enrolment. Moreover, Attanasio and Priulla (2020) showed that student mobility from the South to the Center and North of the country has increased at each level. After the economic crisis in 2008, there was an overall recovery in terms of university enrolments in Italy. However, the authors showed that despite that increase, around 30% of students living in the South decided to enroll in universities in the Center-North in 2017.

Literature has also given insights into the factors associated with mobility choice. The works from Boscaino et al. (2022) and Santelli et al. (2019) highlight southern regions are affected by an increasing rate of students – especially from Sicily – moving to other regions, arguing that better job-market opportunities drive mobility to the North. D'Agostino et al. (2019) and Impicciatore and Tosi (2019) note how the South-North mobility is also affected by contextual factors such as students' social class and family background. The recent work from Genova et al. (2019) highlighted how some preferential paths from Sicily to the Center-North are significant over time, providing evidence that student mobility is not merely a random process from the South to the North of Italy. Student choices are motivated by interpersonal relationships, private information, and strong and weak ties at destinations that compete to shape mobility patterns.

Gender inequalities

In addition to the aforementioned inequalities, the Italian higher education system is, like the majority of institutions globally, marked by gender segregation. The trajectory of young people's education and employment is significantly impacted by gender, which is a fundamental factor in distinction and inequality (Macarie and Moldovan, 2015; Salmieri and Giancola, 2020; Weeden et al., 2020; Barone and Assirelli, 2020). Examining the factors leading to the creation of gender-related inequalities in higher education career paths is crucial for improving equality in education, in the job market, and society as a whole. As previously seen in the other type of inequalities, gender segregation can be described on two levels: the concept of *vertical* gender segregation is used to address female under-representation at higher levels of education, such as master's or doctoral levels; the concept of *horizontal* gender segregation is instead used to address the different educational choices of males and females.

Females outperform males in terms of educational attainment, secondary school, and academic achievements in most European countries nowadays (De Vita and Giancola, 2017; Salmieri and Giancola, 2020). In Italy, females have been more than males in high school attendance since 1981, and the same has been recorded in university attendance since 1991.

According to a sizable body of analysis and research from gender studies, the recent overtaking of females in secondary and tertiary educational attainment has not yet resulted in a reduction of the horizontal segregation in the educational choices made by males and females (Macarie and Moldovan, 2015; Cheryan et al., 2017; Barone et al., 2019; Romito et al., 2020). Research on gender differences in higher education has shown the existence of significant gender inequalities, shaped along the humanistic-scientific divide, with females under-represented in Science, Technology, Engineering, and Mathematics (STEM) or STEM-related fields (Cheryan, 2012; Gabay-Egozi et al., 2015; Tandrayen-Ragoobur and Gokulsing, 2021). These findings are consistent across nations, highlighting the structural reasons that generate gender segregation at different educational levels. On average, across OECD countries in 2020, females made up 31% of new enrollments to the bachelor's level in STEM fields and 79% of new entrants to health and welfare programs in short-cycle tertiary and bachelor's level (OECD, 2022).

Curricular choices in high school are indeed heavily different along gender lines. These can be observed in educational curricula and content preferences, interests and learning orientations, and how teachers, counselors, families, and peer groups approach and condition the academic careers of male and female students (Salmieri and Giancola, 2020). Female students are less likely than male students to attend high school scientific and industrial technical tracks. At the same time, they prefer tracks where more importance is devoted to humanities, relationships between people, and caring for others (Barone et al., 2019). The same trend is clearly reflected in their university field of study choice. Male students are more likely to enroll in STEM programs, which are more expendable in the labor market and heavily linked to future higher returns. Conversely, female students are more interested in humanistic and non-STEM fields in general.

The under-representation of women in STEM careers can be described as a "leaky pipeline". This pipeline carries students from secondary school through university and on to a STEM career in the labor market (Clark Blickenstaff, 2005). Hall and Sandler (1982) offered one reason for this phenomenon, describing STEM fields, particularly engineering ones, as a "chilly climate" for female students. According to the author, scientific departments have higher expectations for male students or give women the impression that their goals are less significant than those of their male coworkers. That attitude might significantly inhibit female student career choices and success in STEM fields. This pipeline leaks students at various stages in their educational career: firstly, in their field of study choice at university, and then along their university careers. The absence of female role models is a further issue. In most industrialized countries, men comprise the majority of scientists and engineers, and while percentages vary from one field to another, the general trend is clear (Clark Blickenstaff, 2005).

However, considering STEM programs as a unique block could be misleading since each program attracts male and female students differently. For instance, female students represent mostly new entrants in some STEM programs, such as biology and life sciences. Conversely, males dominate engineering and computer science programs. Students' stereotypes about the the proportion of men in a field correspond to current gender disparities within STEM, with computer science, engineering, and physics being stereotypically associated with males more than biology, chemistry, and mathematics (Cheryan et al., 2017). Gender inequalities are significant in specific sectors for several reasons. First, fields are missing out on the advantages of gender diversity and the potential contributions of bright females. Second, females may miss out on well-paying and esteemed careers since they are typically associated with the most technical STEM fields.

Some recent studies deal with the gender gap at universities in Italy, with a particular focus on STEM programs. Barone and Assirelli (2020) highlight the crucial role of curricular track choices in Italy, stating that this single factor mediates most of the gender differences in the access to engineering and computer science university programs. Using Italian administrative data, Enea and Attanasio (2020) found that females are more likely than males to graduate with a bachelor's in geology, biology, biotechnology, and statistics. On the contrary, they seem to suffer in all the remaining STEM programs, especially mathematics.

The literature has provided many insights into inequalities in students' choices and performance at university (Belloc et al., 2010; Contini et al., 2018; Aina et al., 2018). However, studies have mainly relied on survey data or small-scale analyses based on particular institutions due to the lack of longitudinal administrative national data archives.

To overcome this problem, a national research project titled "From high school to job placement: micro-data life course analysis of university student mobility and its impact on the Italian North-South divide" started in 2016. It is based on an agreement signed by the Ministry of University and Research (MUR) and the Universities of Palermo, Cagliari, Siena, and Turin and amended in 2017 to include the Universities of Florence, Naples Federico II, and Sassari, and in 2019 the Universities of Cattolica and Enna Kore. This agreement allows the research group to access the ANS national student-level micro-data archives. These databases contain all the information about the students enrolled in Italian universities from 2008 to 2020.

The research group has built an ad hoc database (MOBYSU.IT, 2017) and produced several works concerning the analysis of student university careers (D'Agostino et al., 2019; Genova et al., 2021; Columbu et al., 2021b; Santelli et al., 2022). The MOBYSU.IT database has allowed for overcoming the limitations given by the archives of individual universities. For instance, data from single institutions did not allow tracking of student careers in the case of a university change. This database allows for analyzing the careers of all the students enrolled in Italian universities, following them from first university enrollment to graduation.

The database includes i) data registered at enrolment related to socio-demographic characteristics and high school career of the students, such as the code for the specific high school attended and its location at the municipality level, and the final mark, and ii) data on university performance for each academic year, such as the university program, the average grade on exams, the number of credits collected, etc.

The main objective of this research project was to analyze student mobility from the southern to the central and northern regions of the country. However, the potential of this database allowed for a broader examination of university careers, focusing on dropout rates and degree completion. Moreover, this database allows, for the first time, to learn about the transition from BA graduation (1st level) to MA (2nd level) enrollment, which has been rarely explored due to the lack of longitudinal data. Analyzing this transition is crucial because it deals with the second-level student mobility flows. Also, we gain more information on potential inequalities in university persistence at higher levels of education.

The lack of information about students enrolling in a doctoral program and leaving Italy to attend a foreign university is a drawback of this data. The latter implies that it is impossible to distinguish between students who transfer to a foreign university and those who leave before graduating.

The absence of knowledge regarding socioeconomic background has been overcome thanks to a recent agreement with the National Evaluation Institute for the School System (INVALSI). INVALSI is a research entity with a juridical status that carries out national large-scale standardized computer-based tests. Its main objective is to evaluate the overall quality of the educational system for each different type of school. INVALSI tests are administered annually - save during the COVID pandemic - to students at four levels of education, aiming to evaluate mathematical and Italian language skills and, from 2018, also English reading and listening skills. In addition to the test scores, the INVALSI database includes details on students' high school careers, socioeconomic status, parental educational level, and intermediate high school grades in maths and Italian.

The ANS and INVALSI micro-data have been recently merged. This linkage allows the analysis of student transitions from high school to university in a way that has never been done before. We can reconstruct student careers from their last year of high school using this large integrated database. Currently, this linkage is available only for the fifth-year high school students of the 2018/19 cohort. This means we only have complete information regarding the first-year university performance of the students enrolled in 2019/20. Another drawback of this data is that we do not have information about students enrolling in a foreign university. This means it is impossible to distinguish those students from those who do not enroll at university.

The main aim of this thesis is to investigate the inequalities in the academic outcomes of Italian students. This dissertation is divided into four chapters in an attempt to adhere to a chronological order. Each chapter will deal with a different step in university careers and with a different type of inequality, and different approaches will be used to answer various questions.

In Chapter 1, the ANS-INVALSI database is used to investigate the transition from high school to university. The aim is to estimate the effect of undertaking a relatively new high school curriculum in Italy with more hours devoted to mathematics-related subjects against the traditional track of the scientific "liceo". Multi-level propensity score matching and discrete-time Markov models are used to evaluate the effect of attending the two tracks on two academic outcomes: the choice to enroll at university and first-year performance.

In Chapter 2, an overview of Italian student mobility is provided. The aim is to describe the unidirectional flows from southern to northern regions and provide insights into the differences between *stayers* and *movers* regarding university performance.

Chapter 3 focuses on the study of gender differences in university performance in STEM programs. The aim is to provide detailed insights into how students' first-year university performance helps to predict university graduation. Segmented regression models are used to evaluate the non-linear relationship between first-year performance, intended as the number of first-year credits, and the probability of obtaining a bachelor's within four years, which is the regular time to complete a bachelor's program for the Italian Ministry of Education.

Finally, Chapter 4 focuses on gender inequalities in university student trajectories. In this work, the main university career transitions are analyzed in great detail, from the first enrollment to the bachelor's completion and the subsequent master's degree enrollment. Discrete-time Markov models are used to estimate the probability of persisting up to the enrollment at the master's level based on a set of socio-demographic and high school and university career variables. As far as the author's knowledge, this is the first work in which gender differences are analyzed with such detail, covering the entire university career.

Chapter 1

Does taking additional Maths classes in high school affect academic outcomes?

This chapter is based on the work from Priulla, A., Vittorietti M., & Attanasio, M. (2023). *Does taking additional Maths classes in high school affect academic outcomes?*. Submitted.

Abstract

Several studies from the mathematical education literature show the effect of students' high school math skills on their success at higher levels of education and work. Here, the choice of studying more mathematics in high school depends on many factors related both to the high school system of a country and to students' inclinations and characteristics. As for Italy, taking additional math courses in high school is not random since it depends on several substantial factors. This selection bias implies that the differences in academic outcomes might be traceable not only to ability and knowledge in math. This work aims to estimate the treatment effect of undertaking a relatively new high school curriculum in Italy with more maths against the traditional science track of the scientific "liceo". This is done for two academic outcomes: university enrollment and first-year university performance. After reducing the selection bias using a caliper multi-level propensity score matching procedure, a multi-state Markov model is used for studying the treatment effect on the joint educational outcomes.

1.1 Introduction

Mathematics is widely regarded as one of the most important school subjects and has a central major in high school due to its relevance and application in most fields (Hagan et al., 2020). According to Niss (1994), the reason that societies attach so much importance to mathematics rather than to any other science is ascribable to: i) its essential role in a wide variety of general areas of practice, such as the representation of numbers, the measurement of time, space, weight, and all sorts of graphical representations and tables; ii) its importance for specialized practice areas such as optimization, explanation, prediction, decision-making, and problem-solving; iii) its fundamental role in the formulation and foundation, as well as in the methods and techniques of many other disciplines. The study of mathematics at university underpins the study of many subjects, not only in STEM but also in many other fields (Nicholas et al., 2015). There is at least one mathematics or mathematics-related course in most degree programs, and math courses are often regarded as among the toughest.

In this section, we aim to investigate the effect of studying more mathematics in high school on future academic outcomes. Here, the choice of studying more mathematics in high school depends on many factors related both to the educational system of a country and to students' inclinations and characteristics. In the US, students have more flexibility in choosing courses in high school. The high school system is different in other countries, such as Italy. Italian students are not allowed to choose single courses in high school since they choose, at the age of 13, a five-year high school curriculum among several equal for everybody. Germany's highly stratified educational system, which has an early first transition point, has a similar organizational structure. In the German setting, choosing a course of study following elementary school impacts future academic success. Regarding the possibility of enrolling in postsecondary education, the high school track is more restrictive than the Italian one (Neugebauer et al., 2013).

Generally speaking, students who take advanced maths courses or choose a more math-intensive high school curriculum systematically differ from others in most educational systems (Lee and Ready, 2009; Contini and Scagni, 2013). It has long been known that taking additional maths classes correlates with university enrollment, success (Trusty and Niles, 2003; Poulsen, 2019), and occupational opportunities (Joensen and Nielsen, 2009). Hence, there is a selection bias in the decision to study more mathematics in high school. In the US, selection bias occurs since only the best students decide to study more math in high school (Wang, 2013). In Italy, the choice to study more math, that is, attending a specific high school curriculum, is more related to socio-cultural factors than students' skills and preferences.

Several surveys are conducted to investigate mathematical skills in high school at both the international and national levels. Programme for International Student Assessment (PISA) promoted by OECD is the world's largest and most popular educational survey. In the international framework, PISA tests aim to measure teenagers' learning levels in maths, science and reading. In Italy, there are specific tests managed by INVALSI. These tests, which are, in many respects, similar to PISA ones, are administered yearly to pupils of four different educational levels. They aim to measure mathematical and Italian language abilities and, since 2018, also English reading and listening abilities.

Investigating student performance in high school is of paramount importance since it has proved to be among the most influential factors in university performance (Contini and Scagni, 2013; Ballarino and Panichella, 2016). However, it is known that student university performances depend upon an incredibly intricate net of multi-dimensional factors, including the student's high school career, the socioeconomic status (Checchi and Flabbi, 2007; Barone et al., 2018), gender (Barone, 2011; Contini et al., 2017; Priulla et al., 2021), and geographical differences (Bratti et al., 2007; Agasisti and Vittadini, 2012). However, very little research has been done in Italy to examine how those factors affect academic outcomes outside surveys or small-scale analyses focused on a single institution. This is most likely due to the absence of longitudinal microdata.

In this section, we use the ANS-INVALSI merged database. The linkage between these two large databases allows for unique insights into how factors, such as high school performance and socioeconomic status, influence student academic outcomes. Focusing on Italy, we aim to evaluate the effects of studying more mathematics in high school on two different and crucial academic outcomes: the choice to enroll at university and the first-year university performance.

Similar to a clinical study, two treatments are then considered: the *traditional* track of the scientific "liceo", the most popular high school curriculum in Italy, and the *applied science* track, where more hours are devoted to mathematics-related subjects than at the traditional science track. Petolicchio (2016) compares the in-going and out-going mathematical skills of the students of the two curricula in Italy. The author notes the strengths and the weaknesses of both tracks, but to the best of the authors' knowledge, there are no results about the differences in academic performance between the two scientific tracks.

To assess the effect of attending the applied science track, we first need to consider the previously mentioned selection bias. To do this, we use a propensity score matching procedure to balance the characteristics of the two groups.

After the balancing procedure, a Markov multi-state type model is used to study the treatment effect on the two joint academic outcomes. Those models are typically used in survival analysis for determining the stage progression of the disease: they have rarely been used in an educational framework (Vittorietti et al., 2019). The idea behind this work is that students' choices and performances represent stage transitions. More specifically, we want to estimate the likelihood of having a specific academic performance, expressed in terms of the number of first-year credits, conditioned on enrollment in a specific degree program at a university. These conditional probabilities are calculated for both unmatched and matched datasets. Following the matching procedure, the differences between the two treatment groups are examined to determine the effect of studying more mathematics on university performance.

This chapter is structured as follows: in Section 1.1.1, the structure of the Italian educational system is introduced; in Section 1.2, the data from three Italian administrative sources are described; in Section 1.3 a descriptive analysis is conducted for a first overview of the differences between the two tracks; then, in Section 1.4, the multi-level propensity score matching procedure and the discrete-time multi-state Markov model are briefly set out; finally, the results and conclusions are discussed in Sections 1.5 and 1.6.

1.1.1 Theoretical Framework

Stratification and selectivity describe the structure of the educational system of a country (Lodi, 1982; Ballarino and Panichella, 2021). Regarding school tracking, social scientists intend an educational system highly stratified from a *horizontal perspective*. A school system is said to be horizontally stratified when students are allocated to different, non-overlapping curricula at each stage of their educational career.

The highly stratified Italian educational system allows students to select the best appropriate track from various options. The choice of the type of high school is taken at the age of 13, and it has substantial and long-lasting consequences for future educational career and labor market opportunities.

As previously noticed, many reforms have been proposed aiming to reduce the hierarchy and stratification of the Italian educational system. Nowadays, despite a large variety of tracks, the Italian high school system can still be described as tripartite: "licei", which provide a five-year academic-oriented generalist education (with additional distinctions in humanities, sciences, languages, and pedagogy); five-year technical schools; and five-year vocational schools. Each of these paths is associated with very different outcomes in terms of further education and labor market participation. The "licei" aim at preparing students to complete a university degree (Cappellari, 2004; Panichella and Triventi, 2014). Among all, humanistic and scientific "licei" are the most prestigious and demanding tracks. Technical schools aim to provide students with a scientific and technological background in the economic and technological-professional sectors (Gentili, 2017). In contrast, vocational schools address students to job placement.

After the Second World War, attendance increased for all social classes at each level of education. However, socioeconomic status is still related to the high school curriculum. Using surveys conducted by the Italian National Statistic Institute (ISTAT) on a sample of 20000 high school graduates, Contini and Scagni (2013) analyzed student transition from secondary school to high school and university. The authors found that the socioeconomic background of the family is strongly influential on each educational transition: on the one hand, children from upper classes mainly enroll in humanistic or scientific schools, which prepare students for a possible university enrollment; on the other hand, children from lower classes enroll in technical or vocational schools. This relationship is still present, as shown in Figure 1.1[1], which clearly shows the hierarchical structure of Italian high schools.

As such, choosing a scientific or humanistic curriculum is considered a forward-looking choice and is more likely to result in completing a university degree and obtaining a better job (Triventi, 2013; Panichella and Triventi, 2014). Moreover, Ballarino and Panichella (2016) found out that the effect of the school track is reducing its strength in its mediating role between social origin and university enrollment in recent years. In this sense, students from more advantaged families are far more likely to enroll at university than before.

1.1.2 The scientific "liceo"

Fascism introduced the Gentile reform in 1923. Within this Italian high school system reform, the science "liceo" was initially a four-year course compared to the humanities "liceo", a five-year course. The latter was considered the elite curriculum. Only in 1952

[1] The ESCS index is a measure of a student's socioeconomic status used in both INVALSI and PISA studies. See Section 1.2 for a better index description.

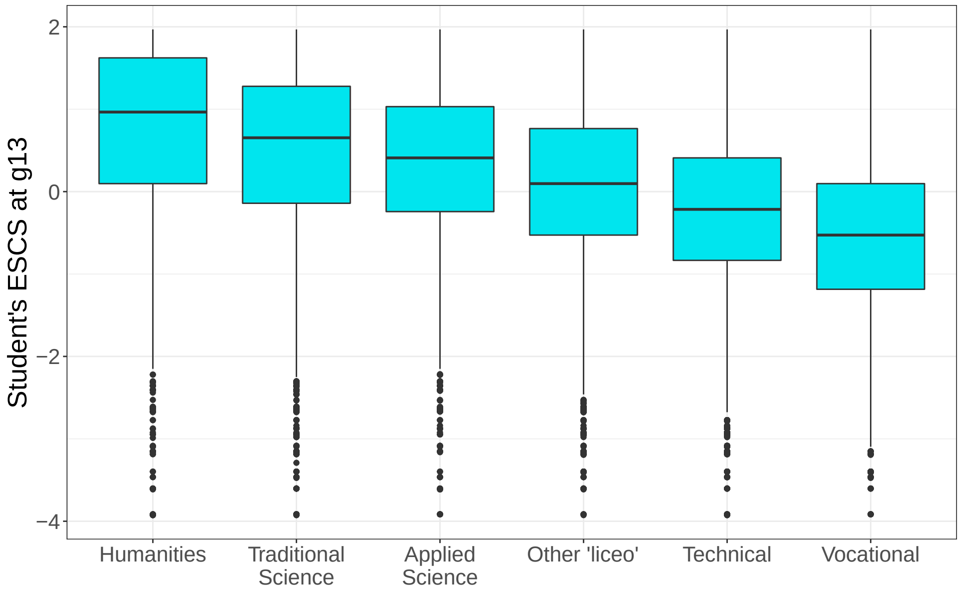

Figure 1.1: Students' ESCS distribution of the students enrolled at the fifth and last year of high school in Italy in 2018/19.

was the science "liceo" transformed into a five-year course. Since 1947, several reforms have been proposed within the Italian high school system. However, its hierarchical structure, as introduced by Gentile, has been slightly attenuated. After the 1969 reform, which allowed university enrollment for all students with a high school diploma, the share of students enrolling at humanistic schools drastically dropped. Conversely, the number of students enrolling at the scientific "liceo" constantly increased, and now it is the most popular high school with a university enrollment rate similar to the humanistic students. Yet even at the traditional scientific high school, where up to 50% of the weekly hours are devoted to literary studies and only 33% to math, physics, and science, the primacy of the humanities is still evident. (Attanasio and Porcu, 2021).

In 2010, the Gelmini reform introduced a new track within the scientific "liceo", the *applied science* track. This new track aimed to provide students with extensive training in studies of scientific and technological knowledge, with a focus on the mathematical, physical, chemical, biological, and earth sciences, as well as computer science and its broad applications. As shown in Table 1.1, the main difference to the traditional science track was that Latin was replaced with more classes devoted to scientific subjects, such as mathematics, computer science, and natural sciences.

Table 1.1: Number of hours per week devoted to scientific and non-scientific subjects in the two tracks of the scientific "liceo" since the 2010 reform.

	Scientific "liceo" track									
	Traditional					Applied Science				
Discipline	**I**	**II**	**III**	**IV**	**V**	**I**	**II**	**III**	**IV**	**V**
Mathematics & Computer Science	5	5	4	4	4	7	6	6	6	6
Physics	2	2	3	3	3	2	2	3	3	3
Natural sciences	2	2	3	3	3	3	4	5	5	5
Scientific	**9**	**9**	**10**	**10**	**10**	**12**	**12**	**14**	**14**	**14**
Italian	4	4	4	4	4	4	4	4	4	4
Latin language	3	3	3	3	3	-	-	-	-	-
Foreign language	3	3	3	3	3	3	3	3	3	3
History and geography	3	3	-	-	-	3	3	-	-	-
History	-	-	2	2	2	-	-	2	2	2
Philosophy	-	-	3	3	3	-	-	2	2	2
Technical drawing and arts	2	2	2	2	2	2	2	2	2	2
Gymnastics	2	2	2	2	2	2	2	2	2	2
Religion	1	1	1	1	1	1	1	1	1	1
Non-scientific	**18**	**18**	**20**	**20**	**20**	**15**	**15**	**16**	**16**	**16**
Total number of hours	**27**	**27**	**30**	**30**	**30**	**27**	**27**	**30**	**30**	**30**

1.2 Data description

The dataset used for the analysis is built by linking data from three distinct administrative national sources:

- **ANS-S**: *Anagrafe Nazionale Studenti - Rilevazione sulle Scuole* was established with the third article of the D.L.gs 76_2005 to collect data for the first and second education cycles in Italy. This allowed MIUR to evaluate the Italian educational system better. This micro-data contains information at the school level, such as the school track (academic, technical, or vocational), a geo-referenced code for the macro-region, region, province, and municipality of the school, the number of students who got their diploma by gender and final mark, and the continuation rate by school track, i.e., the proportion of students who pursue further studies. The available data covers the period from 2014/15 to 2018/19. Thanks to this database, we can investigate the school-university transition in a way that, as far as we know, has never been explored in Italy (Contini and Scagni, 2013; Ballarino and Panichella, 2016).

- **INV-S**: micro-data from the INVALSI. In addition to the test scores, INVALSI also collects information regarding students' profiles. This allows associating several important individual characteristics with academic performance. Among these, additional details such as the socio-demographic status, family background, geographical provenience, or some indicators of past school performance (i.e., whether the student had a regular high school career) are also provided.

- **ANS-U**: micro-level longitudinal data from the National Archive of University Students (ANS). It is a database with all the information about the university careers of all the students enrolled in Italian universities from 2008 to 2020. This database contains a record for each first-year student, including information about their high school background and university career. This data allows us to follow students throughout their university career, from enrollment to dropout or completion.

The linkage of these databases allows investigation of i) the transition from high school to university at school and individual level; ii) the relationship between student performance in high school and university outcomes.

We consider the population of students enrolled in the fifth and final year of high school in 2018/19. In detail, we have information about the choice to enroll at university and first-year performance only for the students enrolled at university in 2019/20 after high school graduation. The students enrolling later at university are then considered not enrolled. This is because it was impossible to link the INVALSI data of the 2018/19 cohort with the ANS-U data of the 2020/21 cohort. Finally, in this work, we consider only those schools that include at least one between the traditional and applied science tracks of the scientific “liceo”.

1.3 Preliminary analysis

This section aims to provide a general framework of the differences between the two scientific tracks. Due to the hierarchical structure of our data, the analysis will be conducted on two different levels: at the track level, we will analyze the distribution of the two tracks across Italy; at the student level, we will consider variables such as the Economic, Social and Cultural Status index (ESCS), the INVALSI test scores, and the transition rates to university.

Here, the ESCS index measures a student’s socioeconomic status. This index, used in both INVALSI and PISA studies, is built based on the following set of variables:

- An indicator of the highest parental occupation, according to the International Socio-Economic Index of Occupational Status. It is computed considering the occupational levels of the father and the mother;

- An indicator of the highest level of parental education converted into years of schooling. It is computed using the International Standard Classification of Education (ISCED) system to be comparable across countries;

- A composite index of family wealth that includes information regarding family possession of educational-related goods, i.e., books, PCs, etc.

Unfortunately, the data at hand does not allow the decomposition of the ESCS index in its three components. By construction, this index has zero mean and standard deviation equal to one (Ricci, 2010). A value of the index above zero indicates that students have a socioeconomic and cultural level below the Italian average and vice-versa.

1.3.1 Track-level analysis

In Figure 1.2, we show the percentage of traditional science or applied science tracks over the total number of tracks offered at the regional level in 2018/19. Due to its recent introduction, the applied science tracks make up only 6.5% of the total number of tracks offered by Italian schools, while the traditional science track represents 17%. The traditional science track is better established in Italy generally, and its distribution on the territory is more homogeneous. The applied science track is mainly established in the northern and central Italian regions, while its presence in the southern regions, especially in the islands, is still negligible. In this respect, the students on the traditional science track are 1.8 times those attending the applied science track in the North, while the mean ratio is around 3.4 in central and southern regions and 4.2 in the Islands.

1.3.2 Student-level analysis

In Table 1.2, we show an overview of the main characteristics of the students attending the two scientific tracks of "liceo". In detail, we are interested in socioeconomic status, gender composition, and high school performance.

The students attending the two tracks present some significant differences: i) students from both tracks have a higher socioeconomic status than the Italian average. On average, applied science students have a lower ESCS index than traditional science ones. This attests to the fact that upper-class families favor the traditional science track; ii) the overall percentage of females is 32% in the applied science tracks and 47.3% in the traditional

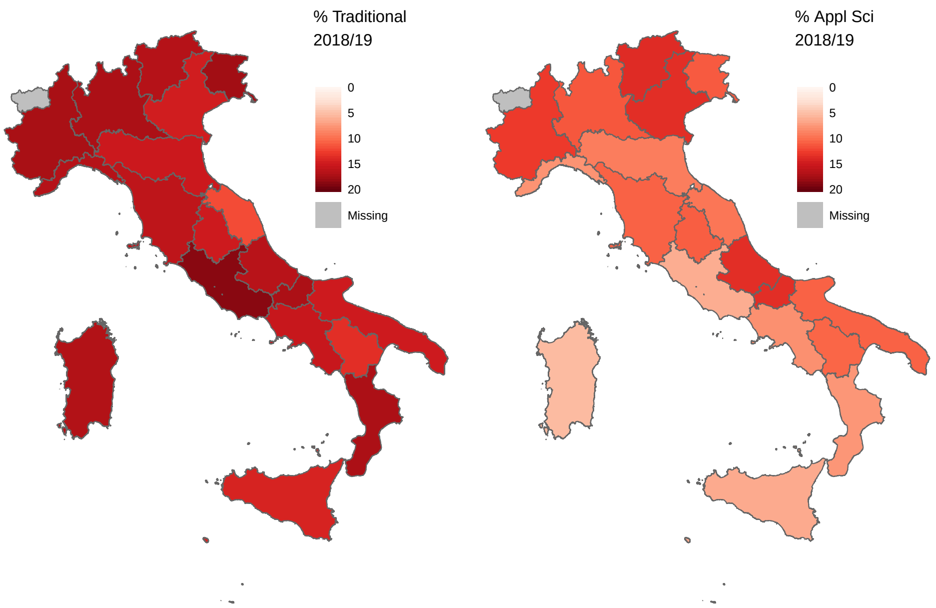

Figure 1.2: Percentages of traditional science or applied science tracks over the total number of tracks offered by high schools at the regional level in 2018/19.

scientific ones. This could be related to the lower humanistic orientation of the applied science track compared to the traditional one; iii) the proportions of students with a non-regular scholastic career and with foreign citizenship, factors usually associated with poorer academic outcomes and lower socioeconomic status, are higher in the applied science track; iv) on average, students from the applied science track perform slightly better in INVALSI maths tests, but slightly worse in the remaining tests.

In Figure 1.3, we show the school-university transition rates for both the traditional and applied science tracks of the scientific "liceo". Here, the transition rate is calculated as the total number of students enrolled at an Italian university in 2019/20 over the total number of fifth-year high school students in 2018/19. It is important to remind that we have no information about students enrolling abroad, which means that they are considered not enrolled. On average, the transition rates are 85.6% and 82% for the traditional and applied science tracks. However, some regional differences can be observed. The transition rates are higher in the northern and central regions for both tracks, but the North-South divide is more evident for the applied science track. This could be related to its more recent introduction and its slower diffusion in southern regions.

Future academic and labor market prospects are clearly influenced by the high school track chosen. In light of this, it makes sense to assume that students from the two tracks,

Table 1.2: Students' characteristics of the two tracks of the scientific "liceo". Cohort of the fifth-year high school students in 2018/19.

	Scientific "liceo" track	
Variable	**Traditional science**	**Applied science**
Student's ESCS		
mean	0.53	0.36
sd	0.93	0.92
Gender		
no. of females	25699	6783
% of females	47.3	32.0
Career regularity		
no. of regular students	51481	19150
% of regular students	94.8	90.2
Citizenship		
no. of foreigners	2429	1307
% of foreigners	4.5	6.2
INVALSI score		
Maths		
mean	237.77	240.08
sd	35.40	34.55
Italian		
mean	226.80	222.41
sd	35.61	33.32
English reading		
mean	227.14	223.98
sd	32.61	32.58
English listening		
mean	222.70	221.30
sd	35.54	35.20
Total	**54327**	**21237**

with different amounts of time devoted to scientific subjects, make different choices concerning the field of study at university. In this regard, in Figure 1.4, we show how the students on the two scientific tracks are distributed according to the university enrollment field of study. The distribution of the students among the different fields seems to be similar, with most students in both groups enrolling in engineering and health degree programs. Yet, there are some observable differences: as for applied science, a higher percentage choose to enroll in computing and engineering programs, while a lower percentage enroll in non-STEM programs, especially in arts and humanities, business and administration, and law.

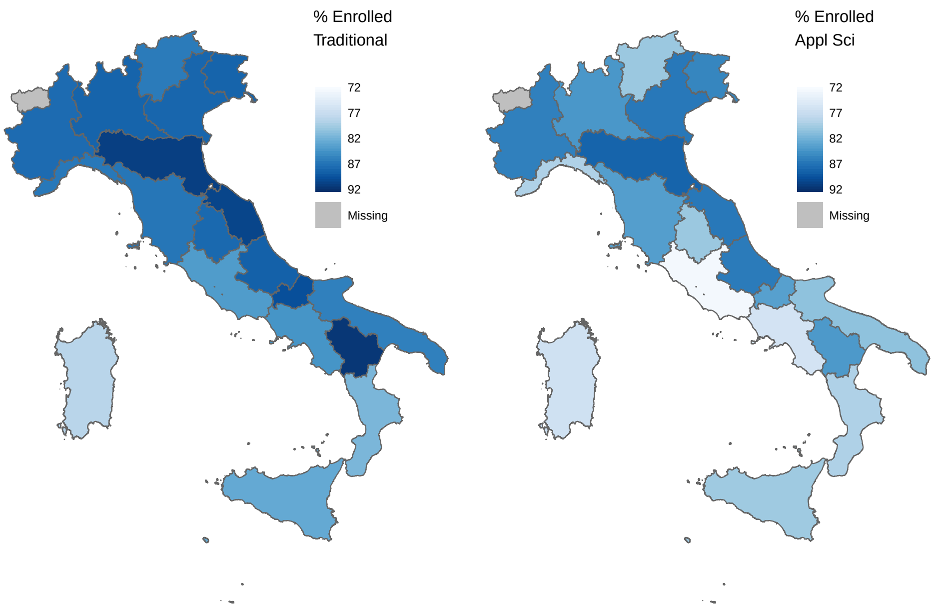

Figure 1.3: School-university transition rates for the two science tracks at the regional level, calculated as the number of students enrolled at university in 2019/20 over the total number of fifth-year high school students enrolled in a specific region in 2018/19.

1.4 Methods

When interventions are randomly assigned, differences between treated and control groups can be interpreted as causal effects. But when subjects select their treatment, different outcomes may reflect initial differences in treated and control groups rather than treatment effects (Cochran and Rubin, 1973). In this context, propensity score matching (PSM) has been widely regarded as a helpful tool i) for reducing the selection bias by balancing the characteristics of the two groups and ii) for estimating treatment effects in observational studies where the treatment assignment is not random (Austin, 2011a). in their seminal work, Rosenbaum and Rubin (1983) defined the propensity score as the conditional probability of a unit being assigned to a treatment given a certain set of covariates. The authors' key insight is that matching merely for the propensity score is enough to eliminate selection bias. Propensity score matching has the advantage of limiting the dimensionality of matching to a single dimension, which greatly aids in the matching process, as opposed to directly matching on the whole set of variables.

In this work, the assignment to the treatment, namely the choice of the high school track, is not random since it depends on many social and cultural factors. Students aged

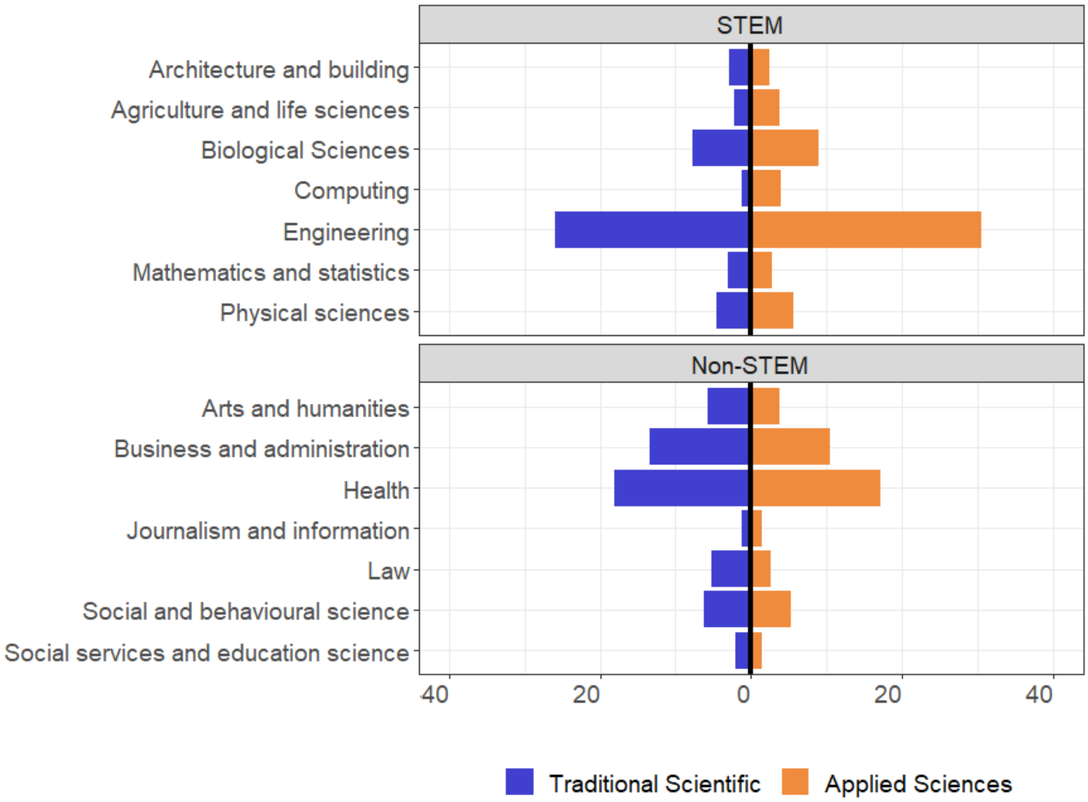

Figure 1.4: Percentage distribution of the students of the two tracks of the scientific "liceo" according to the university enrollment field of study.

13 do not choose the school track by themselves. The parents are usually responsible for their children, and the choice strongly depends on socioeconomic status (Checchi and Flabbi, 2007; Panichella and Triventi, 2014; Ballarino et al., 2014; Giancola et al., 2020), the geographical area in which they live (Bratti et al., 2007), the gender of the children (Contini et al., 2017), and other factors. Therefore, the applied and traditional science treatment groups present substantial differences.

When dealing with hierarchical data, multi-level propensity score matching (MPSM) is the natural extension of the propensity score (Arpino and Mealli, 2011). The multi-level or clustered structure adds a layer of complexity given that the selection mechanism and the dependency within clusters and the outcomes may vary considerably across clusters. Educational data represents a perfect example of a hierarchical structure in which individuals (students) are grouped in clusters (schools) in which they receive a "clustered treatment", namely their high school curriculum (Pimentel et al., 2018). Here, the school "summarizes" the general context regarding the socioeconomic level and the neighborhood in which the schools are located.

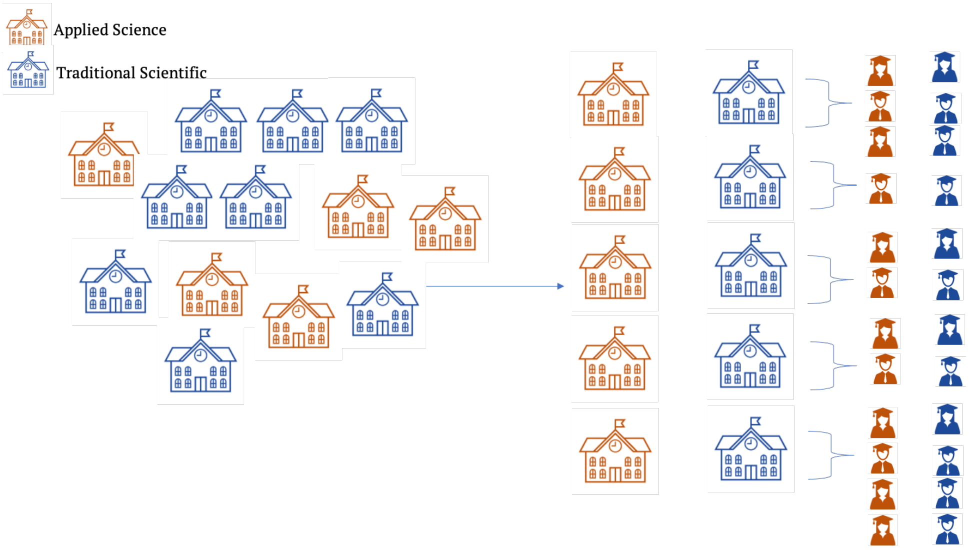

Figure 1.5: Illustration of the two-step matching procedure.

Different methods for handling multi-level educational data have been proposed in the literature (Keele et al., 2021). Arpino and Cannas (2016) reviewed the possible matching procedures for the multi-level context, highlighting the advantages of taking into account the hierarchical structure of the data. However, most literature concerns cases in which the treatment is administered at the individual level. Inspired by the procedure proposed by Rickles and Seltzer (2014), we propose a two-step matching procedure for clustered data in which the treatment is administered at the cluster level. The idea behind this procedure is explained in Figure 1.5: first, we match the tracks using a single-level logit including the set of track-level covariates; then, based on the new dataset of matched tracks obtained in the first step, we match the students using a logit model with random effects given by the paired couples of the two tracks, including the set of student-level covariates.

Once a good balance between the two groups has been achieved, we focus on assessing the effect of attending the two tracks on the two educational outcomes: university enrollment and first-year performance. These two outcomes occur at two different times: first, the enrollment or not enrollment, and second, the first-year performance. Therefore, we decided to respect the longitudinal nature of the data and consider a multi-state type model. In particular, we consider a multi-state model based on a discrete-time Markov chain, also called the Markov chain transitional models, popular in analyzing longitudinal data (Agresti, 2003).

As already seen, students make similar choices concerning the field of study at univer-

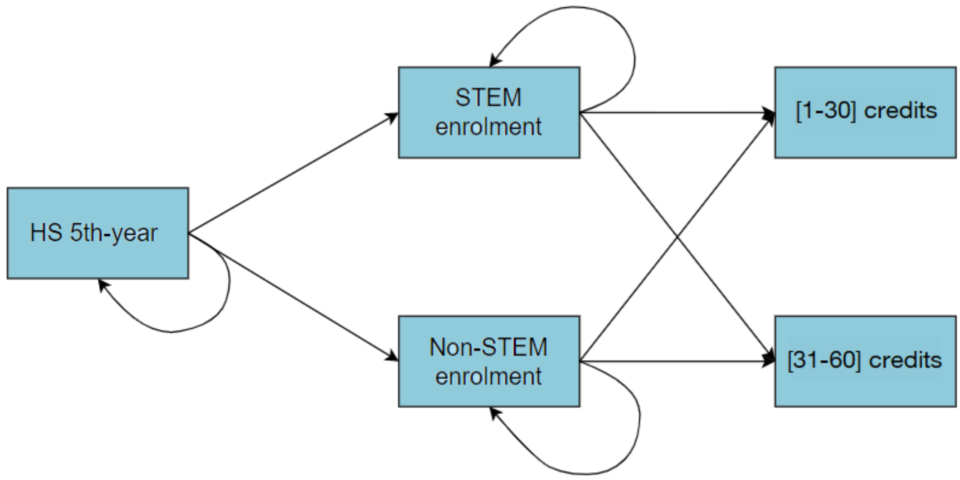

Figure 1.6: States and transitions of the multi-state Markov model.

sity. At the same time, it is known that student performance differs based on the field of study (Attanasio et al., 2018; Ferrão and Almeida, 2019). Hence, we decided to distinguish between enrollment in a STEM or non-STEM program and evaluate student performance in the two areas considering the number of credits earned at the end of the first year, which is among the best predictors for university completion (Attanasio et al., 2013). In this framework, we can imagine that each career step is a state that potentially occurs between the fifth year of high school and the end of the first year of university. Using this representation, it is reasonable to assume a Markov chain where each student has a different transition probability from each state to another based on their profile. In Figure 1.6, the possible transitions are shown: at $t = 0$, students are in the fifth year of high school; at $t = 1$, the students can decide to enroll in a STEM or non-STEM program or not enroll at university at all, which means that they remain in the initial state; at $t = 2$, two transitions are possible based on the number of credits obtained at the end of the first year: ≤ 30 and > 30, as a rough classification of "bad" and "good" performing students.

1.4.1 Multi-level propensity score matching

This section outlines the structure of the multi-level propensity score matching procedure.

Consider a two-level data structure where N first-level units (students), indexed by i ($i = 1, 2, \ldots, n_j$), are nested in J second-level units (tracks), indexed by j ($j = 1, 2, \ldots, J$). We consider a binary treatment T_{ij} for the first-level unit administered at the track level, such that $T_{ij} = 1$ if cluster j is treated and $T_{ij} = 0$ otherwise. Each first-level unit has two potential outcomes Y_{ij}: $Y_{ij}(1)$ under treatment condition, $Y_{ij}(0)$ under control condition. Let $\boldsymbol{Z}$ and $\boldsymbol{X}$ be the matrices of first- and second-level covariates.

The propensity score is defined as:

$$\pi_{ij} = \pi_{ij}(\boldsymbol{X}, \boldsymbol{Z}) = P(T_{ij} = 1 \mid \boldsymbol{X}, \boldsymbol{Z}). \tag{1.1}$$

Usually, in observational studies, the propensity score must be estimated from the data. To this end, propensity score estimates are typically obtained with either logistic fixed- or random-effects models. Here, we propose a procedure to deal with the multi-level structure of our data which works as follows:

1. A single-level logit model is fitted to estimate the propensity score at the cluster level:

   $$g(\pi_j) = g(P(T_j = 1 \mid \boldsymbol{Z})) = \boldsymbol{\gamma_0} + \boldsymbol{\gamma}_{\boldsymbol{Z}}^T \boldsymbol{Z}, \tag{1.2}$$

   where $\boldsymbol{\gamma_0}$ is the vector of intercept, $\boldsymbol{\gamma_Z}$ is the vector of the cluster covariates effects and g denotes the logit link function.

   The most common PSM methods are greedy and optimal matching (Austin, 2011a). Despite the superior performance of optimal matching in terms of balance, this method has been regarded as computationally expensive and troublesome. Greedy matching, instead, can sometimes be too simplistic as an approach and lead to poorly-balanced groups. Introducing a caliper in the greedy matching approach is a good compromise between the classical greedy matching and the optimal matching approach. Considering a caliper in propensity score matching, we allow matching among units within a chosen threshold of the propensity score (Arpino and Mealli, 2011).

   Formally, let I_1 and I_0 denote, respectively, the set of treated and control units, and let A_r indicate the set of control clusters matched to the treated cluster $r \in I_1$:

   $$A_r = \{k \in I_0 : \hat{\pi}_k = \min_{k \in I_0} \mid \hat{\pi}_r - \hat{\pi}_k \mid < c_1\}, \tag{1.3}$$

   where c_1 is the caliper imposed on the cluster-level covariates. After (1.3) has been constructed for all clusters in the treatment group, the matched dataset M is built:

   $$M = \{r : A_r \neq \emptyset\} \cup \left\{ \bigcup_r A_r \right\}.$$

2. A random-effects logit model is fitted to estimate the propensity score at the individual level, considering the couples obtained in the previous step and belonging to the matched dataset M as random effects:

$$g(\pi_{ir}) = g(P(T_{ir} = 1 \mid \boldsymbol{X})) = \boldsymbol{\alpha} + \boldsymbol{\beta}_{\boldsymbol{X}}^T \boldsymbol{X}, \boldsymbol{\alpha} \sim N(\alpha_0, \sigma_\alpha^2) \tag{1.4}$$

where $\boldsymbol{\alpha}$ is the vector of random effects and $\boldsymbol{\beta}_{\boldsymbol{X}}$ is the vector of individual-level covariates effects. As in the first step, we use a one-to-one caliper matching at the individual level based on estimated propensity scores (1.4).

In this procedure, both the first- and second-level calipers are set to 0.05, based on the work of Austin (2011b). To assess the goodness of the matching procedure, we use the difference in means, the most commonly chosen method for exact matching (Ho et al., 2007). The resulting matched dataset will include the subset of students whose characteristics are better balanced at the school and student levels.

It is important to stress that we decided to use this specific matching procedure because it was the most rational choice given the aim of our work. However, other matching procedures were performed with no relevant differences from the selected procedure.

The matching procedure has been carried out using the R package `MatchIt` (Stuart et al., 2011).

1.4.2 The discrete-time multi-state Markov model

The class of multi-state models is an extension of the competing-risks models. Those models deal with one initial state and several mutually exclusive absorbing states. Multi-state models are helpful when the individual's process also consists of intermediate events that cannot be classified as initial or final states (Putter et al., 2007).

A multi-state model is a model for a stochastic process that occupies one of a set of discrete states at any time. These classes of models are usually used, for example, in medical research to analyze a patient's disease or recovery process. In that case, states can describe a patient's medical condition, like healthy, diseased, and dead, while a state change is a transition usually corresponding to a disease outbreak, complications, and death.

It is known that pinpointing an event's exact moment of occurrence when working with longitudinal educational data is a bit impossible. In this case, the problem does not arise since students are observed in three distinct moments of their educational career that are equal for everyone: i) the enrollment in the fifth year of high school, ii) the potential enrollment at university after completing high school, and iii) the end of the first academic year.

Generally, given S = 1,. . . ,R a discrete set of states, a stochastic process $\{M_n\}$ is called a Markov chain if for all times $n \geq 0$ and all states $i_0, \ldots, i, j \in S$,

$$P(M_{n+1} = j \mid M_n = i, M_{n-1} = i_{n-1}, \ldots, M_0 = i_0) =$$
$$= P(M_{n+1} = j \mid M_n = i) = P_{ij}.$$

P_{ij}, often referred to as *transition probability*, denotes the probability that the chain, whenever in state i, moves to state j. Roughly speaking, $\{M_n\}$ is characterized by the so-called Markov property that states that the probability of being in a given state at a given time depends only on the state occupied in the immediately previous time and not on the whole process history. P_{ij} are then entries of the so-called *transition probability matrix*, $\mathbf{P}$, that in our case can be represented as:

$$\begin{array}{cc} & \begin{array}{ccccc} \text{Not enr.} & \text{STEM} & \text{Non-STEM} & \text{[1-30] credits} & \text{[31-60] credits} \end{array} \\ \mathbf{P} = & \begin{bmatrix} P_{11} & P_{12} & P_{13} & 0 & 0 \\ 0 & P_{22} & 0 & P_{24} & P_{25} \\ 0 & 0 & P_{33} & P_{34} & P_{35} \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{array}{l} \text{Not enr.} \\ \text{STEM} \\ \text{Non-STEM} \\ \text{[1-30] credits} \\ \text{[31-60] credits} \end{array} \end{array} \tag{1.5}$$

The non-zero entries of the matrix 4.3 identify the possible transitions. The rows of $\mathbf{P}$ satisfy the condition $\sum_{j=1}^{5} = 1$. We assume *time homogeneity* for the Markov chain. Hence, the transition probabilities do not depend on the time n. In a discrete-time multi-state Markov model, covariates are usually linked to the transition probabilities p_ij through a set of multinomial logistic regressions. More specifically, a multinomial logistic regression is fitted for each row of the transition probability matrix:

$$log\Big(\frac{P_{ij}}{P_{ii}}\Big) = b_{ij,0} + b_{ij}^T X, \; i \neq j$$

.

Two discrete-time multi-state Markov models are then fitted: the first model is

$$log\Big(\frac{P_{ij}}{P_{ii}}\Big) = b_{ij,0} + b_{ij}^T X_{TREATMENT},\ i \neq j, \tag{1.6}$$

where $X_{TREATMENT}$ is the only covariate considered that identifies the two treatments: traditional and applied science tracks; The second model is

$$log\Big(\frac{P_{ij}}{P_{ii}}\Big) = b_{ij,0} + b_{ij}^T X_{TREATMENT} * X_{MACROREGION},\ i \neq j, \tag{1.7}$$

where we introduce the $X_{MACROREGION}$ covariate because the preliminary results showed the higher diffusion of the applied science in the northern regions (Figure 1.2), and difference in macroregional enrolment rates between the two tracks (Figure 1.3).

The Markov models are performed using the R package `msm` (Jackson, 2007).

1.5 Results

In this section, the results of the MPSM procedure and the multi-state Markov models are shown.

First, the results of the MPSM procedure are shown in Figure 1.7. This procedure provides a good balance of the characteristics of the two groups. The imbalance in the overall percentage of females favoring the traditional science track at the track level has been almost totally removed. In addition, a good balance has been reached concerning the macro-regional location of the two tracks. The only factor for which the two groups are still slightly out of balance is the school's ESCS.

At the student level, the student's ESCS has been almost entirely balanced, as well as gender, career regularity, and citizenship. A good balance is also obtained for the INVALSI score in mathematics, originally higher for applied science students. Instead, the whole balancing procedure significantly worsens the other INVALSI subjects' scores.

Then, the multi-state Markov model results for the model in Equation 1.6 are reported in Table 1.3. The model is fitted on the matched and unmatched datasets to observe the true treatment effect after balancing. Overall, the differences between the two tracks in estimated transition probabilities can be observed to have shrunk after the balancing procedure. Yet, some differences persist. The results of the model fitted on the matched data highlight that students from the applied science track have a lower probability of

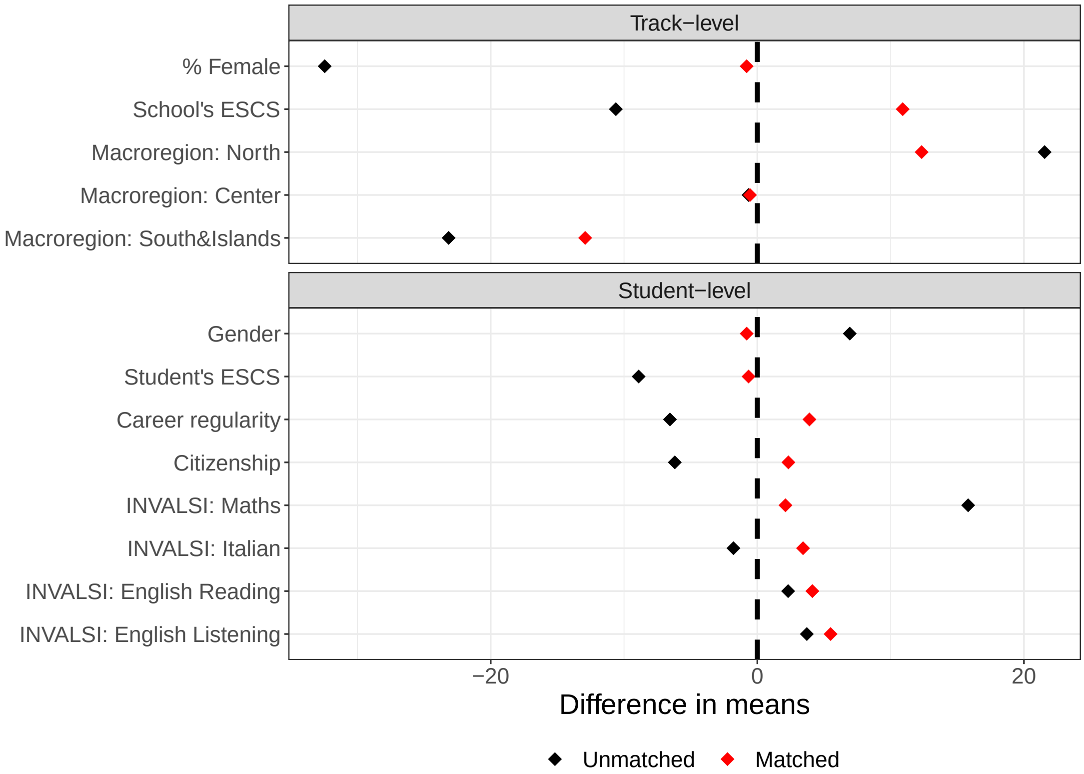

Figure 1.7: Difference in means between the two tracks of the scientific "liceo" of the track- and student-level covariates before (red) and after (black) the MPSM procedure. Positive values indicate an imbalance in favor of the applied science track, and negative values in favor of the traditional science track.

enrolling at university than those from the traditional science track. On the one hand, students who attended the applied science track have a higher probability of enrolling in a STEM program than their traditional science peers; on the other hand, the probability of enrolling in non-STEM programs is reversed. As for academic performance, before the matching procedure, the probability of remaining in the state STEM, namely the probability of not obtaining any credits after having enrolled in STEM, was significantly higher for the applied science students, but this difference is no longer significant after the matching procedure. As before the matching, for STEM students, the probabilities of moving to the non-zero credits states were significantly higher for students from the traditional science track. Those differences are, meanwhile, no more significant after the matching. There are, at that point, no more differences between the two tracks in transition probabilities from non-STEM to any state. The only exception concerns the probability of transitioning to more than 30 credits, which is significantly lower for applied science students in non-STEM programs.

The estimated probabilities of the second model are then shown in Figures 1.8, 1.9,

Table 1.3: Estimated transition probabilities for both tracks of the scientific "liceo" before and after the MPSM procedure. CIs are in brackets.

	Unmatched	**Matched**
Transition	**Traditional science**	
HS last-year -> HS last-year	0.142 (0.138,0.145)	0.159 (0.153,0.166)
HS last-year -> STEM	0.425 (0.421,0.430)	0.432 (0.424,0.441)
HS last-year -> Non-STEM	0.434 (0.429,0.438)	0.408 (0.399,0.418)
STEM -> STEM	0.060 (0.057,0.064)	0.070 (0.064,0.077)
STEM -> (0-30] credits	0.257 (0.251,0.262)	0.271 (0.259,0.284)
STEM -> (30-60] credits	0.683 (0.677,0.689)	0.658 (0.645,0.671)
Non-STEM -> Non-STEM	0.055 (0.052,0.058)	0.071 (0.064,0.078)
Non-STEM -> (0-30] credits	0.189 (0.184,0.194)	0.194 (0.183,0.206)
Non-STEM -> (30-60] credits	0.756 (0.750,0.762)	0.735 (0.723,0.748)
	Applied Science	
HS fifth-year -> HS fifth-year	0.180 (0.175,0.185)	0.181 (0.174,0.188)
HS fifth-year -> STEM	0.487 (0.481,0.494)	0.477 (0.467,0.486)
HS fifth-year -> Non-STEM	0.332 (0.326,0.340)	0.342 (0.334,0.351)
STEM -> STEM	0.075 (0.070,0.080)	0.077 (0.070.0.084)
STEM -> (0-30] credits	0.280 (0.271,0.289)	0.278 (0.267,0.291)
STEM -> (30-60] credits	0.644 (0.636,0.654)	0.644 (0.631,0.657)
Non-STEM -> Non-STEM	0.079 (0.073,0.086)	0.081 (0.072,0.089)
Non-STEM -> (0-30] credits	0.211 (0.202,0.221)	0.215 (0.202,0.229)
Non-STEM -> (30-60] credits	0.709 (0.698,0.720)	0.704 (0.689,0.718)

1.10, and 1.11. In Figure 1.8, we showed the estimated probabilities for the HS last year $\rightarrow$ Not enrolled transition. Macroregional differences between the two tracks are evident. The estimated unmatched probabilities of not enrolling at university: i) between the two tracks are larger in the Center and South; ii) increase from North to South, especially for the applied science track. On the other hand, the estimated matched probabilities: i) do not change after the matching procedure for the applied science track; ii) increase for the traditional science track. In detail, the difference between the two tracks is no more significant in the North, while it is less pronounced in central and southern regions.

In Figure 1.9, the estimated probabilities for the *HS last-year* $\rightarrow$ *STEM* and *HS last-year* $\rightarrow$ *Non-STEM* transitions are reported. The results show that the matching procedure: i) does not change the probabilities for the applied science track; ii) determines a decrease in the probabilities of enrolling in non-STEM (especially in the Center) for the traditional science track; iii) determines a slight increase in the probability of enrolling in STEM for the traditional track.

Figure 1.10 shows the estimated probabilities for the *STEM* $\rightarrow$ *0 credits* and *Non-STEM* $\rightarrow$ *0 credits* transitions. The results show that: i) southern students are more likely

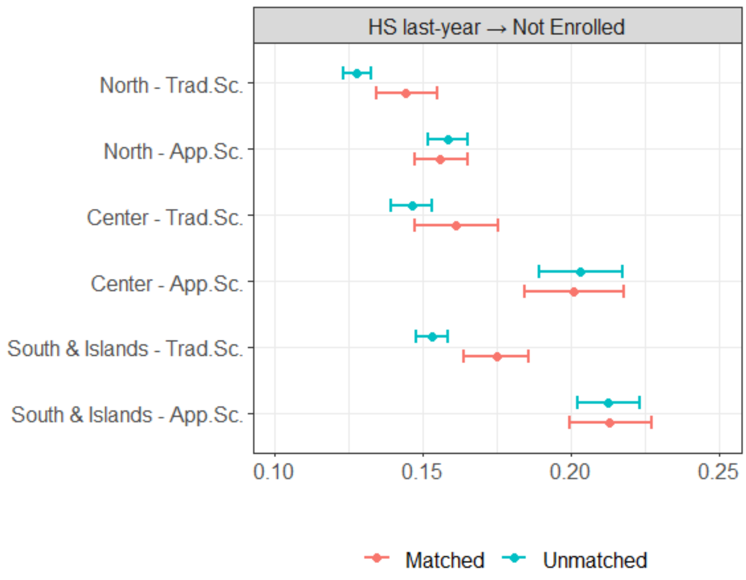

Figure 1.8: Estimated probabilities for the *HS last-year → Not enrolled* transition by macroregion of the school, before and after the matching procedure.

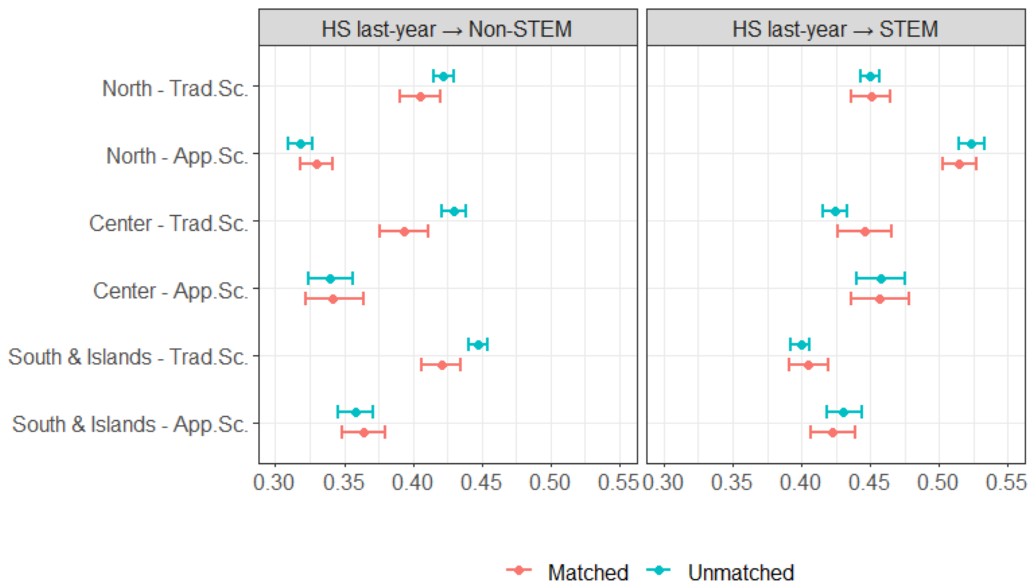

Figure 1.9: Estimated probabilities for the *HS last-year → STEM* and *HS last-year → Non-STEM* transitions by macroregion of the school, before and after the matching procedure.

not to achieve credits in the first year; ii) after the matching procedure, students from the central high schools perform the worst in STEM programs, while southern students perform the worst in non-STEM ones; iii) the disparity favoring the traditional scientific track is more pronounced in southern regions in non-STEM programs, while it is more evident in central regions in STEM programs.

Finally, figure 1.11 shows the estimated probabilities for the *STEM → (30-60] credits*

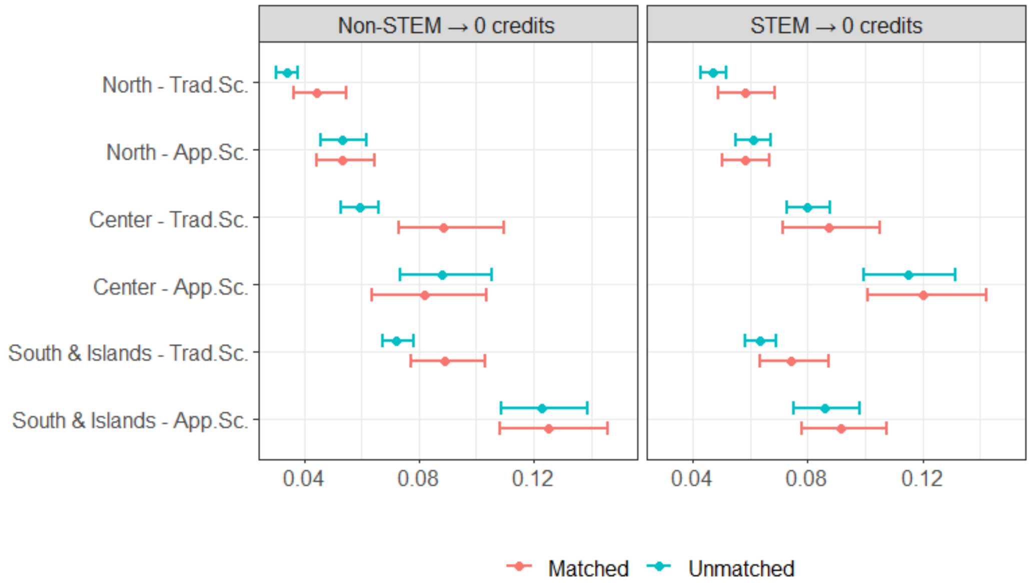

Figure 1.10: Estimated probabilities for the *STEM → 0 credits* and *Non-STEM → 0 credits* transitions by macroregion of the school, before and after the matching procedure.

and *Non-STEM → (30-60] credits* transitions. The results show that: i) northern students are more likely to achieve more than 30 credits in the first year, in both STEM and non-STEM programs; ii) the matching procedure determines, despite some exceptions, a decrease in these probabilities; iii) the disparity between the two tracks is significant for northern and southern non-STEM students, and for central and southern STEM students.

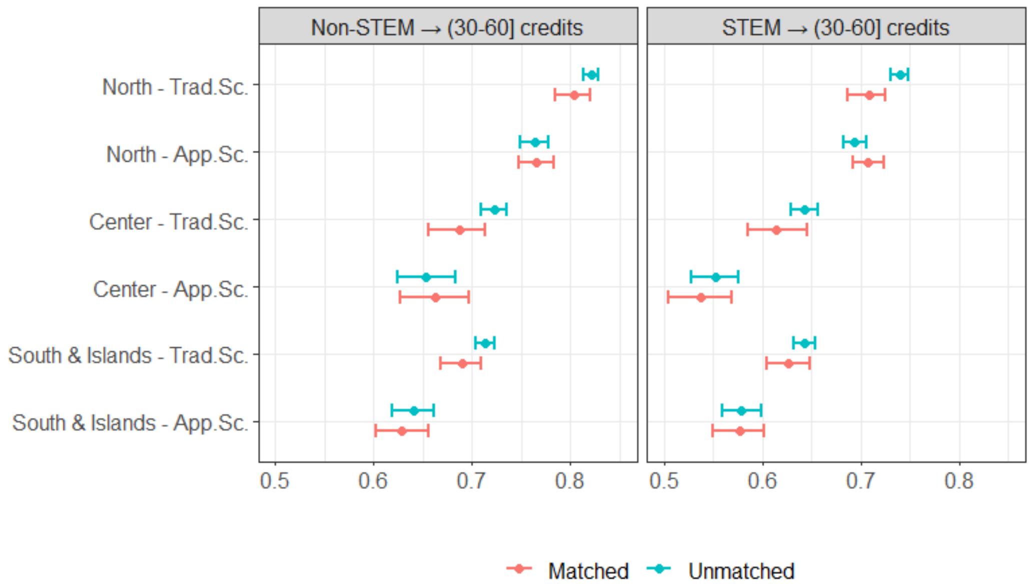

Figure 1.11: Estimated probabilities for the *STEM → (30-60] credits* and *Non-STEM → (30-60] credits* transitions by macroregion of the school, before and after the matching procedure.

1.6 Conclusions

Studying the effect of mathematics on student academic outcomes is not a trivial matter. Academic outcomes, such as university enrollment and first-year university performance, depend on an incredibly intricate net of multi-dimensional factors related to students' characteristics and school outcomes. These include socioeconomic status, gender, and the type of high school track. As a proxy for the role of math on student careers, we considered the high school track attended. We focused on the two tracks of the scientific "liceo": the traditional and the applied science tracks, the latter offering more hours devoted to mathematics and mathematics-related subjects. However, the students from the two tracks mentioned above have many other substantial differences than just the hours given over to mathematics. Hence, due to the hierarchical structure of our data, we used multi-level propensity score matching to create two balanced groups of students. This procedure was necessary because of the imbalance between the two tracks regarding gender composition, socioeconomic status, and high school student performance.

Then, we used a discrete-time multi-state Markov model to examine the effects of studying more math on university enrollment and first-year performance in terms of credits. Our results highlight that, after having balanced the two groups, the gap between the two tracks of the scientific "liceo": i) has been reduced in terms of the transition from high school to university; ii) has almost disappeared in terms of university performance. The estimated transition probabilities show that students from the applied science track tend to be less likely to enroll at university than their traditional science peers. Studying more math is a push factor for university enrollment in more scientific programs, as students on the applied science track tend to enroll more in STEM programs than those on the traditional science track. The higher interest of applied science students in STEM programs could be considered a natural career choice. As for university performance, the only worth remark is that traditional science students enrolled in non-STEM programs have a higher probability of obtaining more than 30 credits than applied science ones. This is probably because of the more humanistic-learning nature of the traditional science track. Moreover, the macroregional analysis shed light on the disparities in the Italian territory. Northern students are generally more likely to enroll at university and perform better than central and southern students in STEM and non-STEM programs. Yet, beyond this evidence, we are aware that our results suggest it is hard to entirely disentangle the hierarchical structure of the Italian high school system, in which the traditional science "liceo" occupies a better rank compared to the applied science “liceo”. This is due to several unobservable variables, likely given by the expectations of students and teachers and the general educational atmosphere.

Finally, as a more general consideration, the traditional science track, though it has fewer scientific classes than the applied science track, is still seen as being “better”. This reflects the still important idea that humanities education gives a wider-ranging education than a purely scientific one.

Chapter 2

Analysis of university student mobility from the South to the Center-North of Italy

This chapter is based on the work from Attanasio, M., & Priulla, A. (2020). *Chi rimane e chi se ne va? Un'analisi statistica della mobilità universitaria dal Mezzogiorno d'Italia.* In: "Verso Nord. Le nuove e vecchie rotte delle migrazioni universitarie". Franco Angeli Editore.

2.1 Introduction

In 1999, the "Bologna Process" proposed the realization of the European Higher Education Area (EHEA) space to promote knowledge, mobility, and cultural cohesion between European countries and the rest of the world. These objectives were reiterated in the Bucharest Communiqué (2012), approved by the Ministers of Higher Education of the 47 EHEA members, in which they emphasized the need to ensure, on the one hand, the highest possible level of public funding for higher education as a form of investment for overcoming the financial crisis and, on the other hand, equal access to higher education in Europe. The latter aspect is relevant to our country, which is not only not competitive at the European level due to low investment in research and development but is also plagued by a South-North territorial dualism in terms of employment and education (SVIMEZ, 2014).

In addition, the Italian university system has undergone reforms that have produced many changes over the past 25 years. The financial autonomy of universities, introduced with Law 537 of 1993 as a tool to contain expenses, accompanied by increasing cuts in public spending, has triggered a mechanism of competition among universities and conditioned education policies to the logic of the market. "The fact that the rewards mechanisms inherent in university funding have diverted resources from the South to the Center-North is puzzling. It is true that, on average, southern universities are less efficient. Still, the cure cannot pass through a diet of their funding, resulting in an impoverishment of human capital and an acceleration of youth mobility to the rest of the country and abroad" (Livi Bacci, 2015). This policy has widened the North-South gap regarding services offered, research quality, and student attractiveness, giving a boost to student mobility in the transition from high school to university. Vast is the literature on this topic (Bratti et al., 2008; Dotti et al., 2013). Today, conversely, we are faced with a rapidly aging South and increasingly becoming a "supplier of skilled human resources to the rest of the country" (SVIMEZ, 2014).

Further evidence that the roots of these selective mobility patterns can be found in students' inter-regional mobility comes from the fact that southern students who have attended a northern university have very little intention of returning to the South. According to SVIMEZ (2014), about 25% of southern students attend a university in the North or Centre of the nation, and only one-third of those students return to the South after graduating. The other two-thirds remain in the North and Centre.

In Italy, university mobility has been studied using the ANS administrative data from MUR or individual data from individual universities' archives on different spatial scales. These studies show that student mobility depends not only on local universities but also on local labor market conditions in the origin and destination areas (Dal Bianco et al., 2010; Dotti et al., 2013). However, data used in these works did not allow for longitudinal analysis of student careers at the national level.

In this chapter, the aim is to briefly describe mobility flows from the South to the central and northern regions of the country over the past decade. This is done to draw the geography and recent history of the regions and universities with the largest outflows and the regions and universities with the greatest attractiveness. In this regard, we define *mover* as a student moving to a region different from his/her region of residence for study-related reasons and *stayer* as a student who remains in the same region. Two moments of the university career will be explored: the first enrollment at university and the transition to the master's level after bachelor's graduation. The latter has been rarely investigated due to the lack of longitudinal data covering the whole university career.

As for the first university enrollment, the data considered in this chapter are restricted to the university careers of students enrolled for the first time in an Italian university in the academic years 2008-09 and 2017-18. As for the master's degree enrollment, we consider those students enrolled in the academic years 2008-09 and 2014-15.

2.2 Student mobility in Italy at university enrollment

This section analyzes student mobility at enrollment in 3-year programs of Italian universities in the years 2008 and 2017. Initially, we provide an overview of regional mobility, describing flows from the region of residence to the region of university enrollment, controlling for some socio-demographic and high school variables. Then, the focus moves to the outflows from the South to the main centers of attraction represented by the central and northern Italian universities.

2.2.1 Inter-regional flows

Figure 2.1 shows the regional outflows and inflows for the 2008 and 2017 cohorts. It can be noticed that the outflows from southern regions, which were already of concern in 2008, had a noticeable increase in 2017: students leaving Sicily have more than doubled, rising from about 2800 to more than 6500. Apulia shares a similar behavior, although the outflow was already significant in 2008 (more than 5000 students), more than in other regions. Campania is different since outflows are stable and incoming flows increase. In contrast, despite an increase in outgoing mobility in the North, Lombardy and Emilia-Romagna remain two of the largest attraction poles, with more than 10000 students coming from other regions in 2017. Piedmont also follows this trend, in fact, in 2017, the number of incoming students increased by about 2000 units. This increase is mostly due to the recently increasing attraction exerted by the Polytechnic University of Turin on southern students.

Mobility at enrollment is investigated using the origin-destination (OD) matrices. This matrix is frequently used to describe mobility patterns. Here, the origin is represented by the region of residence, while the destination is the region where the university of enrollment is located. In Tables 2.1 and 2.2, the last columns report three rates: the

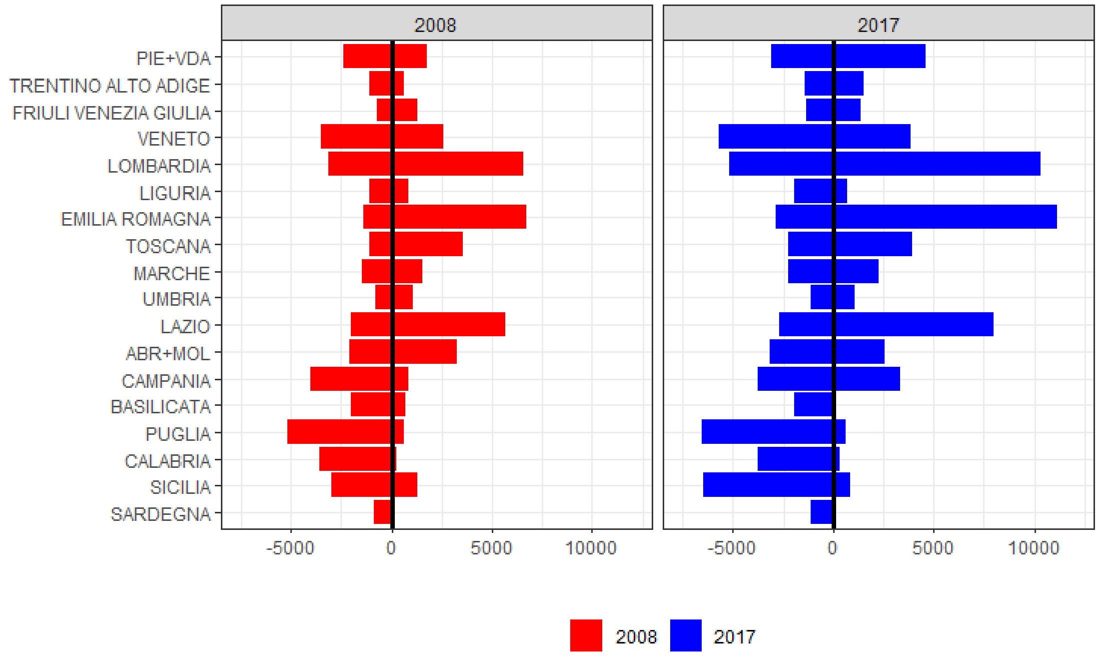

Figure 2.1: Inflows and outflows at enrollment in 3-year degree programs of Italian universities by region of residence. University students enrolled in Italy in 2008/09 and 2017/18.

percentage of *stayers* (%ST), the mobility rate (MR), and the adjusted mobility rate (AMR). Both mobility rates are defined later in this section. The marginal row totals OD(i) are the enrolled students living in the $i-th$ region. Conversely, the marginal column totals OD(j) are the students enrolled in the $j-th$ region. The non-diagonal elements OD($i, j; i \neq j$) correspond to *movers*, that is, the number of students who graduated from high school in the $i-th$ region and who enrolled at a university in the $j-th$ region. Finally, the diagonal elements OD(i, i) are the *stayers*.

The two mobility rates are defined as follows:

- MR: unadjusted mobility rate, calculated as 1 minus the ratio of the number of stayer students over the total number of students living in a region.

$$MR(i) = \left(1 - \frac{OD(i,i)}{OD(i)}\right) * 100 \tag{2.1}$$

- AMR: adjusted mobility rate, calculated as one minus the ratio of the sum of the stayers plus the movers to a neighboring region and the total number of students living in a region.

$$AMR(i) = \left(1 - \frac{OD(i,i) + \sum_{i=1}^{I} OD(i,j)}{OD(i)}\right) * 100 \tag{2.2}$$

where $i \neq j$, with j being neighboring regions.

The AMR considers the flows directed to the neighboring regions, which are not considered movers in calculating the index. Thus, for example, a move from Tuscany to Emilia-Romagna is not considered mobility. It is also pointed out that, although they are not two neighboring regions, Sicily and Calabria have been considered as such, given the big amount of commuters in the Strait of Messina.

From the OD matrices, we see mobility increasing throughout Italy over the last decade, with a decrease of around 6 points in the percentage of stayer students. Yet some regional differences come clear. As expected, the rate of stayers is higher in central and northern regions, with Lazio, Tuscany, Emilia-Romagna, and Lombardy exceeding 90% in 2008. In contrast, the decline in the rate of stayers is more pronounced in the South. In fact, the rate in Sicily decreased from 86% to 68.3% in 2017. The same happens in Apulia, which shows a sharp increase in outgoing mobility. Both southern regions have been significantly affected by the influence of the large poles of attraction in the North in recent years, such as the Polytechnic University of Turin and the University of Bologna, which attract an increasing number of students from southern regions. In contrast, Basilicata is the region with the highest mobility rate, even considering the AMR correction: this is due to the lack of university in the region, which pushes many students to pursue their studies elsewhere in Italy.

Data dramatically shows the unidirectional mobility from the South to the Center-North of Italy. The significant flow of students leaving the South is directed toward the Center-North of Italy, and, among these, Lombardy and Emilia-Romagna appear to be the most attractive regions. Nonetheless, Campania seems to be the only one among southern regions to succeed in "retaining" its students, with a percentage of stayers around 86% for both cohorts. On the other hand, outflows from central and northern regions are never directed to southern regions.

2.2.2 The high school background

As previously noticed, high school background strongly influences educational choices and performance in Italy, as it can be considered a proxy for socioeconomic status. In this regard, it is reasonable to imagine that students from scientific and humanistic schools

Table 2.1: Origin-destination (OD) matrix from the region of residence to the region of first university enrollment. University students enrolled in Italy in 2008/09.

Region of residence	Region of first university enrollment																				
	ABR+MOL	BAS	CAL	CAM	EMI	FRI	LAZ	LIG	LOM	MAR	PIE+VDA	PUG	SAR	SIC	TOS	TRE	UMB	VEN	TOT	%ST	MR	AMR
ABR+MOL	6087	1	-	56	292	4	706	2	218	446	45	18	-	3	164	2	66	39	**8149**	74,7	25,3	15,7
BAS	145	1096	69	222	246	2	317	3	162	35	67	430	-	1	258	2	46	12	**3113**	35,2	64,8	48,8
CAL	86	27	6686	95	312	7	804	20	323	36	129	71	-	1176	365	3	116	39	**10295**	64,9	35,1	23,5
CAM	734	598	60	24111	242	21	1393	22	303	74	87	40	1	11	338	9	100	36	**28180**	85,6	14,4	10,0
EMI	16	2	3	24	12928	25	73	17	643	250	20	6	2	10	101	8	16	155	**14299**	90,4	9,6	6,7
FRI	11	-	-	10	52	3740	28	1	127	10	13	3	-	5	27	6	1	397	**4431**	84,4	15,6	10,4
LAZ	810	7	4	222	83	7	22781	14	151	95	44	15	8	15	256	3	245	23	**24783**	91,9	8,1	3,6
LIG	2	-	1	5	171	3	37	4596	266	5	228	3	3	6	373	1	3	17	**5720**	80,3	19,7	15,8
LOM	42	4	12	34	1411	28	80	82	31657	38	282	19	2	30	117	96	15	812	**34761**	91,1	8,9	3,4
MAR	184	-	2	8	594	11	213	8	190	4892	18	3	-	-	83	3	110	57	**6376**	76,7	23,3	17,0
PIE+VDA	13	-	7	14	94	18	39	490	1514	17	13147	5	1	14	92	7	10	52	**15534**	84,6	15,4	2,7
PUG	1060	104	12	103	1007	38	838	22	752	278	337	14012	1	5	399	11	102	109	**19190**	73,0	27,0	25,9
SAR	14	-	-	4	130	6	114	32	180	21	139	-	5527	4	157	4	17	48	**6397**	86,4	13,6	13,6
SIC	63	-	91	37	551	28	395	52	706	48	258	10	4	18322	574	7	45	125	**21316**	86,0	14,0	13,8
TOS	20	-	2	13	332	6	203	82	171	29	28	4	3	7	12275	4	152	56	**13387**	91,7	8,3	3,9
TRE	3	1	-	1	200	32	15	7	192	10	10	6	1	6	53	2227	5	576	**3345**	66,6	33,4	22,3
UMB	36	-	-	4	57	4	352	10	73	67	3	1	-	2	144	2	2684	21	**3460**	77,6	22,4	8,7
VEN	26	-	-	10	931	1086	81	12	639	44	79	10	1	9	75	495	10	14251	**17759**	80,2	19,8	9,7
TOT	**9352**	**1840**	**6949**	**24973**	**19633**	**5066**	**28469**	**5472**	**38267**	**6395**	**14934**	**14656**	**5554**	**19626**	**15851**	**2890**	**3743**	**16825**	**240495**	**83,6**	**16,4**	**10,8**

Table 2.2: Origin-destination (OD) matrix from the region of residence to the region of first university enrollment. University students enrolled in Italy in 2017/18.

Region of residence	Region of first university enrollment																				
	ABR+MOL	BAS	CAL	CAM	EMI	FRI	LAZ	LIG	LOM	MAR	PIE+VDA	PUG	SAR	SIC	TOS	TRE	UMB	VEN	TOT	%ST	MR	AMR
ABR+MOL	4442	-	-	176	752	22	606	7	325	747	211	13	1	-	160	18	54	95	**7629**	58,2	41,8	31,1
BAS	145	723	8	292	282	6	235	2	159	78	137	390	-	-	179	4	48	18	**2706**	26,7	73,3	49,7
CAL	78	14	4968	315	420	19	688	18	457	52	243	73	3	816	368	5	121	55	**8713**	57,0	43,0	33,0
CAM	662	69	10	23218	364	19	1436	9	500	68	139	47	2	9	315	16	49	55	**26987**	86,0	14,0	10,5
EMI	21	-	157	82	14736	31	265	14	1140	427	101	6	1	1	159	70	16	384	**17611**	83,7	16,3	7,8
FRI	6	-	1	23	160	3366	96	2	222	19	41	1	-	-	34	50	5	679	**4705**	71,5	28,5	15,9
LAZ	458	2	43	454	218	25	25089	14	600	74	175	9	2	7	285	24	247	68	**27794**	90,3	9,7	6,5
LIG	2	-	-	27	202	10	171	4197	472	6	522	8	-	1	456	9	5	34	**6122**	68,6	31,4	11,1
LOM	37	-	10	207	1801	38	676	57	34861	28	810	8	3	13	142	261	15	1055	**40022**	87,1	12,9	6,4
MAR	196	-	1	23	941	27	187	4	358	4817	82	2	2	1	125	26	136	111	**7039**	68,4	31,6	23,4
PIE+VDA	16	-	3	131	177	21	268	344	1920	14	14637	1	4	6	63	24	6	98	**17733**	82,5	17,5	7,6
PUG	778	84	20	454	1367	48	881	9	1012	431	776	11925	1	6	392	77	74	151	**18486**	64,5	35,5	32,7
SAR	10	1	-	38	173	22	147	41	196	30	205	6	4933	3	158	12	14	65	**6054**	81,5	18,5	18,5
SIC	93	1	58	751	1062	83	1080	39	1104	98	921	36	5	13867	764	37	84	212	**20295**	68,3	31,7	31,6
TOS	16	-	1	121	643	9	417	76	486	42	87	4	-	2	12397	20	200	110	**14631**	84,7	15,3	9,1
TRE	5	-	2	29	356	27	59	5	227	10	30	2	-	1	44	1675	7	609	**3088**	54,2	45,8	18,7
UMB	20	-	1	22	151	8	457	3	131	77	53	3	-	-	168	12	2504	38	**3648**	68,6	31,4	12,1
VEN	17	-	2	169	2049	961	353	10	1012	42	91	7	4	6	103	886	14	14518	**20244**	71,7	28,3	17,5
TOT	**7002**	**894**	**5285**	**26532**	**25854**	**4742**	**33111**	**4851**	**45182**	**7060**	**19261**	**12541**	**4961**	**14739**	**16312**	**3226**	**3599**	**18355**	**253507**	**77,7**	**22,3**	**15,4**

could be more likely to enroll in a university located in another region. Then, the relationship between the high school background and the mobility choice is investigated in Table 2.3. In this instance, mobility is determined on a macro-regional scale. This means that, for instance, a student moving from Veneto to Piedmont will be considered a mover. Students from technical and vocational schools have been merged into a single category, as they displayed a similar mobility pattern.

It can be seen that the growth in the mobility rates recorded in 2017 is higher for students who attended scientific (from 11.8% to 15.6%) and humanistic (from 15.8% to 21.7%) "licei". At the macro-regional level, as expected, students from the South show the highest mobility rates, especially for those coming from the humanistic (31.7%) and scientific (25.3%) "licei". The largest increase is recorded for the Islands, mostly due to Sicilian students: again, it is scientific or humanistic graduates, with percentages of 27.8% and 30.9%, respectively, in 2017.

Table 2.3: AMR by macro-region of residence and type of high school attended. University students enrolled in 2008/09 and 2017/18.

	Sci. Liceo		Hum. Liceo		Tec-Voc Inst.		Other liceo		Total	
Macro-region	**%Mov**	**Tot**	**%Mov**	**Tot**	**%Mov**	**Tot**	**%Mov**	**Tot**	**%Mov**	**Tot**
2008										
North	6,7	33066	11,5	8137	5,5	34445	6,6	20201	**6,7**	**95849**
Center	6,3	16444	7,7	6766	4,7	16012	5,6	8784	**5,8**	**48006**
South	20,6	25650	26,8	7879	15,6	25097	16,2	10301	**18,8**	**68927**
Islands	15,5	9411	16,4	4145	12,5	9269	10,4	4888	**13,7**	**27713**
Total	**11,8**	**84571**	**15,8**	**26927**	**9,1**	**84823**	**9,0**	**44174**	**10,8**	**240495**
2017										
North	8,2	34517	14,8	7508	5,9	37782	8,3	26764	**7,9**	**106571**
Center	10,5	17453	13,4	5759	5,5	14788	7,7	13101	**8,6**	**51101**
South	25,3	24294	31,7	6810	19,4	15676	16,5	14855	**22,4**	**61635**
Islands	27,8	8857	30,9	3433	18,4	7298	21,6	5230	**24,2**	**24818**
Total	**15,6**	**85121**	**21,7**	**23510**	**9,8**	**75544**	**11,4**	**59950**	**13,4**	**244125**

2.2.3 Central-northern universities: which are the most attractive?

Outgoing mobility from the South is now considered in detail in Figure 2.2, focusing on universities located in the North-Central regions with the largest attractiveness. Specifically, we consider the top eight universities based on the number of incoming students from southern regions: in 2008, those received about 52% of incoming flows from the southern regions, while in 2017 the percentage dropped to 45%.

Despite the recent decline, La Sapienza attracts the largest number of southern students, especially from Campania and Sicilia. In 2017, the largest increases in incoming flows were recorded at the Polytechnic of Turin and the University of Bologna. In this respect, Sicily is the only region whose outgoing flows have increased, excluding those directed to Bocconi. In detail, the major pole of attraction for Sicilian students is the Polytechnic of Turin, which in recent years has significantly increased its incoming flows from Apulia as well. Sicilians and Apulians also made up most of the southern students enrolling at the Universities of Bologna and Parma. Campania, which, as noted earlier, has the lowest mobility rate among the southern regions, lost most of its students (about 580 in 2008, 400 in 2017) in favor of La Sapienza, located in a neighboring region. Students from Campania are then less attracted to the universities in the North, except for the relatively small outflows directed to the University of Bologna.

Figure 2.2: Enrollment inflows from southern regions in the top eight attractive universities. University students enrolled in Italy in 2008/09 and 2017/18.

2.3 Bachelor's degree achievement: who is faster?

In this section, university success is analyzed from two different perspectives: does mobility have a positive effect on bachelor's degree attainment? How do students from the South perform compared to those from the central and northern regions? In addition, a focus is made on the type of high school attended. To analyze the bachelor's degree completion, the 2017 cohort is replaced with the 2014 one due to the unavailability of complete data related to three-year graduates from the first cohort. Students enrolled at online universities are excluded.

It is important to consider that the difference between stayers and movers should be read with caution because movers are reasonably the most motivated students and those with greater economic means. The costs associated with the decision to study in another region are indeed not negligible, so mover students are more motivated to complete the bachelor's faster than others.

In Table 2.4, the BA graduation rates within four years of southern students are reported, distinguishing between stayers and movers. The difference between the two groups is noticeable: in fact, the percentage of students who graduate within four years is significantly higher for mover students. The largest difference is observed in the Islands, where the BA graduation rate is 57.3% for movers and 41.2% for stayers in 2014. As for the South, the difference in favor of movers is 14 percentage points for both cohorts. Calabrian students perform the worst, with BA graduation rates of 35.1% and 46.3%, respectively, for stayers and movers in 2014. On the other hand, the highest BA completion rates are observed for the students from Campania and Apulia. However, the difference between the two groups is still significant and close to 16% on average in favor of movers for both regions.

2.3.1 The high school background

In this section, the analysis of the university success of southern students is carried out according to their mobility choice and the type of high school attended in Table 2.5.

At the macro-regional level, the BA graduation rates increased for Islands stayers students by 9.3 percentage points. As expected, findings show that students from humanistic and scientific "licei" perform better at university. On the other hand, those with a technical or vocational diploma show a lower and declining probability of bachelor's graduation.

Table 2.4: BA graduation rates within four years by region of residence and mobility at enrollment. University students enrolled in Italy in 2008/09 and 2014/15.

	2008				2014			
	Stayers		Movers		Stayers		Movers	
Region of residence	%BA	TOT	%BA	TOT	%BA	TOT	%BA	TOT
ABR+MOL	30,6	6087	48,7	2062	43,0	4343	56,6	2480
BASILICATA	31,3	1096	42,7	2017	35,4	698	52,8	1754
CALABRIA	28,8	6686	36,0	3609	35,1	4988	46,3	2764
CAMPANIA	28,2	24111	43,6	4069	42,9	20622	59,6	3423
APULIA	30,9	14012	47,0	5178	44,1	10768	60,5	5218
SOUTH	**29,3**	**51992**	**43,5**	**16935**	**42,1**	**41419**	**56,3**	**15639**
SARDINIA	28,6	5527	43,3	870	40,6	4506	50,5	1030
SICILY	22,5	18322	45,6	2994	41,4	12019	58,8	4672
ISLANDS	**23,9**	**23849**	**45,1**	**3864**	**41,2**	**16525**	**57,3**	**5702**
TOT	**27,6**	**75841**	**43,8**	**20799**	**41,9**	**57944**	**56,6**	**21341**

An overall increase in BA graduation rates in 2014 and a narrowing gap between movers and stayers are also worth noting. Nonetheless, the difference in BA graduation rates remains in favor of movers: in detail, the gap decreased from 15.5 to 11.5 percentage points for students with humanistic or scientific backgrounds in the South, and from 22.2 to 13 percentage points for stayers in the Islands with the same background.

Table 2.5: BA graduation rates within four years by macro-region of residence and type of high school attended. University students enrolled in Italy in 2008/09 and 2014/15.

		2008							
		Hum/Sci liceo		Tec-Voc high school		Other liceo		Total	
Macro-region	Mobility	%BA	TOT	%BA	TOT	%BA	TOT	%BA	TOT
South	**Stayers**	36,7	25769	33,6	23865	30,8	8575	34,6	58209
	Movers	52,2	7000	49,6	5595	44,2	1378	50,4	13973
Islands	**Stayers**	30,3	11298	27,0	9405	24,2	4340	28,0	25043
	Movers	52,5	2062	47,2	1601	44,0	475	49,5	4138
Total	**Stayers**	**34,8**	**37067**	**31,7**	**33270**	**28,6**	**12915**	**32,6**	**83252**
	Movers	**52,3**	**9062**	**49,1**	**7196**	**44,1**	**1853**	**50,2**	**18111**
		2014							
South	**Stayers**	48,9	23576	30,1	13272	44,0	7404	42,4	44252
	Movers	60,4	8125	40,8	2434	56,8	1496	56,0	12055
Islands	**Stayers**	46,1	9145	27,2	4523	43,4	2805	40,5	16473
	Movers	59,1	3490	40,0	1191	56,1	766	54,5	5447
Total	**Stayers**	**48,1**	**32721**	**29,4**	**17795**	**43,8**	**10209**	**41,9**	**60725**
	Movers	**60,0**	**11615**	**40,6**	**3625**	**56,5**	**2262**	**55,5**	**17502**

2.4 Mobility at master's degree enrollment

The focus moves to the study of student mobility in the transition to the master's level. In this regard, mobility is defined as enrollment at a university different from the one where the bachelor's degree was earned. This step is more significant than the transition from high school to university since students should be more informed about the choice because of the experience gained after the bachelor's degree. The outgoing flows from southern universities observed in the first transition come up again in the transition to the master's level, contributing again to the impoverishment of southern regions.

In this section, we briefly consider master's enrollment rates within five years after first university enrollment, according to socio-demographic characteristics and university career. The focus will again be restricted to the outflows from the South since, as in the school-BA transition, the BA-MA transition is confirmed to be unidirectional. As in the previous section, the cohorts examined regard students enrolled in bachelor's degree programs at Italian universities in the academic years 2008/09 and 2014/15. Finally, we consider only students who graduated within four years to make the MA enrollment rates of the two cohorts comparable.

2.4.1 Inter-regional flows

In this section, we describe the inter-regional flows at master's enrollment in Italy. We construct, as before, origin-destination matrices in which the origin is the $i-th$ region of bachelor's graduation and the destination is the $j-th$ region master's level enrollment. In Tables 2.7 and 2.8, three rates of interest are computed. First, %MA(i) indicates the percentage of students enrolling at the master's level in the $i-th$ region. This rate is obtained as

$$\%MA(i) = \left(1 - \frac{OD(i, DROP)}{OD(i)}\right) * 100. \tag{2.3}$$

wheree OD(i) are the bachelor's graduates in the $i-th$ region, and OD($i, DROP$) represents the number of students who, after obtaining a bachelor's within four years, decided not to enroll at the master's level within five years. Then, in the last column, the percentage of stayers is computed as

$$\%ST(i) = \left(\frac{OD(i,i)}{OD(i) - OD(i, DROP)}\right) * 100 \tag{2.4}$$

where OD(i, i) are the elements of the main diagonal, thus the students enrolling in the same university where they completed the bachelor's. Finally, the regional attractiveness rate is reported in the last row. This rate indicates the number of students enrolled in a specific region after graduating with a bachelor's in another region. This is obtained as

$$\%Attractiveness(j) = \left(1 - \frac{OD(i,i)}{OD(j)}\right) * 100 \qquad (2.5)$$

where OD(j) is the total number of master's enrollments in the $j - th$ region.

Before moving to the OD matrices, outgoing flows from the South to the Center-North at BA enrollment and MA enrollment are briefly compared in Table 2.6. It can be seen that student mobility in the BA-MA transition is a phenomenon with a greater magnitude than the first one. Moreover, this phenomenon has increased its strength in the last few years. The difference between the two rates for the 2008 cohort is minimal in the South. Conversely, this gap is larger in the Islands: in fact, there has been an increase of 11.5 and 8.6 percentage points, respectively, in the mobility rates at BA and MA enrollment in 2014.

Table 2.6: Mobility rates from the South to the Central-North regions in the school-BA and BA-MA transitions. Absolute values and percentages. University students enrolled in Italy in 2008/09 and 2014/15.

		BA enrollment		MA enrollment	
BA macro-region	**Value**	**2008**	**2014**	**2011/12***	**2017/18***
South	n	11806	12069	1650	3125
	%Mov	17,1	21,1	16,7	25,8
Islands	n	3637	5460	964	1447
	%Mov	13,1	24,6	29,5	33,2
Total	n	**15443**	**17529**	**2614**	**4572**
	%Mov	**16,0**	**22,1**	**19,9**	**27,8**

Note: the academic years 2011/12 and 2012/13 (2017/18 and 2018/19) correspond to the years of BA-MA transition for the cohorts of students enrolled in 2008/09 (2014/15).

The OD matrices show that the average master's level enrollment rate is higher in the Northern regions. In contrast, the highest rates in the South are recorded in Campania and Sicily, with around 65% in 2014. Despite the slight increase between the two cohorts in the master's enrollment rate, mobility remains a huge problem. The rate of stayers decreased from 86.4% to 80.3% in 2014. The only exception is Basilicata, where the rate increased from 28.7% to 56.2%. Conversely, it fell from 77.3% to 61.3% in Abruzzo and

Molise and from 65% to 54.8% in Sicily.

The attractiveness index highlights how the interest of Italian students falls more on northern regions, such as Emilia-Romagna and Piedmont, which in 2014 received about 30% of the flows from other regions, especially from the South. Although the overall mobility rate decreased in 2014, some central and northern regions have shown a substantial increase in attractiveness in 2014: the rise for Emilia-Romagna was 13.3 percentage points, followed by Piedmont and Tuscany, which showed an increase of 9.9 and 7.9 percentage points, respectively.

2.4.2 The attractiveness of northern universities

In this section, the analysis of student mobility is being conducted from a different angle in Tables 2.9 and 2.10, i.e., that of the universities. The objective of this analysis is to identify the universities in the South losing more students in the transition to the master's level and, at the same time, individuate the North-Central universities with the largest inflows from southern universities. In addition, we report the master's enrollment rates at the major southern universities, defining a stayer as a student who remains in the same university after bachelor's graduation.

The OD matrices show that, excluding Naples Federico II, the mobility rates ranged from 68% to 79% in southern universities and from 57% to 69% in the island universities in 2008. These rates decreased, on average, by 9 points in the South and 4 points in the Islands in 2014. The only university that does not follow the trend of increasing mobility is the University of Palermo, whose rate of stayers increased by 6.3 percentage points. Looking at the OD matrices below, it is possible to observe interesting origin-destination couples: Turin Polytechnic plays the role of a strong attraction pole for graduates from the Universities of Palermo and Catania. Similarly, the flows from the University of Bari toward the first three major attraction centers significantly grew in 2014.

Figure 2.3 shows the outflows from the ten largest universities in southern Italy for the 2008 and 2014 cohorts. The "losses" of southern universities and the consequent inflows of northern universities in the BA-MA transition have grown over the past decade. The inflows and outflows reflect again the unidirectional mobility affecting southern universities. While, as expected, the outflows are mainly directed to northern and central regions, incoming mobility is almost entirely due to inflows from other southern regions. Among the southern universities, the University of Bari shows the largest increase in outflows,

Table 2.7: Origin-destination (OD) matrix from the region of bachelor's degree within four years to the region of master's enrollment within five years. University students enrolled in Italy in 2008/09.

	Master's enrollment region																				
BA region	**ABR+MOL**	**BAS**	**CAL**	**CAM**	**EMI**	**FRI**	**LAZ**	**LIG**	**LOMB**	**MAR**	**PIE+VDA**	**PUG**	**SAR**	**SIC**	**TOS**	**TRE**	**UMB**	**VEN**	**DROP**	**TOT**	**%MA**	**%ST**
ABR+MOL	1185	-	-	15	59	2	81	-	36	36	40	10	-	1	20	2	7	39	1575	**3108**	49,3	77,3
BAS	4	72	-	109	9	-	21	-	7	-	1	19	-	-	7	-	-	2	287	**538**	46,7	28,7
CAL	-	-	807	12	59	6	47	-	43	2	30	3	-	18	32	1	8	6	1007	**2081**	51,6	75,1
CAM	17	1	1	3935	68	9	173	5	102	20	58	8	-	1	46	1	4	29	2484	**6962**	64,3	87,9
EMI	11	-	-	3	4549	23	77	5	366	34	125	12	1	6	52	37	7	159	4092	**9559**	57,2	83,2
FRI	1	-	-	1	55	943	18	3	47	7	21	1	-	-	17	8	-	129	1036	**2287**	54,7	75,4
LAZ	59	-	3	29	62	3	6056	8	166	22	55	13	3	5	55	6	8	59	4771	**11383**	58,1	91,6
LIG	-	-	-	-	12	5	3	1046	100	-	46	-	1	-	22	-	-	9	939	**2183**	57,0	84,1
LOMB	1	-	-	6	181	11	99	24	11282	6	117	7	-	2	41	33	-	101	8088	**19999**	59,6	94,7
MAR	13	-	-	5	79	3	55	-	59	1276	54	4	-	1	20	11	6	38	1227	**2851**	57,0	78,6
PIE+VDA	2	-	-	-	29	3	10	18	106	7	3763	-	-	1	12	6	1	26	2653	**6637**	60,0	94,5
PUG	22	-	-	3	104	9	100	1	124	16	84	1972	-	-	43	-	1	50	1902	**4431**	57,1	78,0
SAR	-	-	-	1	35	10	23	4	48	10	54	-	602	-	38	2	1	25	723	**1576**	54,1	70,6
SIC	4	-	4	5	124	9	86	9	156	10	165	4	-	1679	75	3	2	75	1987	**4397**	54,8	69,7
TOS	4	-	1	6	125	12	59	9	100	6	75	4	1	2	2653	9	7	45	2768	**5886**	53,0	85,1
TRE	-	-	-	1	31	14	3	1	25	-	14	-	-	-	7	510	1	54	911	**1572**	42,0	77,2
UMB	9	-	-	6	60	4	57	1	35	8	26	2	-	-	21	1	621	17	790	**1658**	52,4	71,5
VEN	-	-	-	2	149	62	25	4	154	4	33	-	-	2	21	71	3	4203	3691	**8424**	56,2	88,8
TOT	**1332**	**73**	**816**	**4139**	**5790**	**1128**	**6993**	**1138**	**12956**	**1464**	**4761**	**2059**	**608**	**1718**	**3182**	**701**	**677**	**5066**	**40931**	**95532**	**57,2**	**86,4**
Attractiveness	11,0	1,4	1,1	4,9	21,4	16,4	13,4	8,1	12,9	12,8	21,0	4,2	1,0	2,3	16,6	27,2	8,3	17,0	-	-	-	-

Table 2.8: Origin-destination (OD) matrix from the region of bachelor's degree within four years to the region of master's enrollment within five years. University students enrolled in Italy in 2014/15.

	Master's enrollment region																				
BA region	**ABR+MOL**	**BAS**	**CAL**	**CAM**	**EMI**	**FRI**	**LAZ**	**LIG**	**LOMB**	**MAR**	**PIE+VDA**	**PUG**	**SAR**	**SIC**	**TOS**	**TRE**	**UMB**	**VEN**	**DROP**	**TOT**	**%MA**	**%ST**
ABR+MOL	1166	-	-	21	183	12	180	-	80	53	74	18	-	2	45	7	12	50	1459	**3367**	56,7	61,3
BAS	-	100	-	9	23	1	12	-	4	1	7	7	-	-	5	2	2	5	119	**298**	60,1	56,2
CAL	5	-	755	9	73	2	61	-	64	6	41	4	-	21	34	-	16	13	671	**1779**	62,3	68,4
CAM	19	1	1	4776	172	18	299	1	231	17	163	6	-	-	80	3	8	59	3254	**9130**	64,4	81,6
EMI	4	-	-	5	5274	35	114	8	453	66	193	6	2	7	130	60	8	300	5039	**11749**	57,1	79,1
FRI	-	-	-	1	81	776	17	4	55	1	39	1	-	-	26	14	3	154	1073	**2252**	52,4	66,2
LAZ	22	1	6	38	198	13	7059	7	293	25	105	7	-	11	66	34	21	118	5183	**13234**	60,8	88,0
LIG	-	-	-	1	51	7	14	1031	164	2	68	1	-	1	16	12	-	15	898	**2282**	60,6	74,5
LOMB	4	-	-	6	394	29	106	23	12468	13	190	3	1	4	71	43	3	242	9954	**23609**	57,8	91,7
MAR	18	-	-	8	284	25	85	1	126	1375	80	17	-	8	70	12	23	72	1279	**3503**	63,5	62,4
PIE+VDA	-	-	-	4	82	12	33	17	229	2	3966	3	2	2	26	12	4	45	3315	**7764**	57,3	89,3
PUG	28	-	-	10	244	12	124	-	229	38	168	2019	-	1	71	24	11	82	1849	**4937**	62,5	66,0
SAR	-	-	-	1	93	10	29	12	76	7	78	1	637	-	41	4	2	54	756	**1806**	58,1	61,0
SIC	6	-	10	12	251	26	102	9	215	6	250	3	-	2241	66	16	9	91	1793	**5127**	65,0	67,6
TOS	2	-	4	8	218	16	106	11	231	5	126	3	2	5	2649	21	21	73	2849	**6367**	55,3	75,7
TRE	-	-	-	-	93	7	22	-	78	3	44	-	-	-	16	630	2	148	1023	**2078**	50,8	60,4
UMB	4	-	-	1	80	4	45	-	37	12	26	1	-	4	26	5	643	23	719	**1631**	55,9	70,6
VEN	1	-	-	3	287	81	48	3	266	9	121	3	2	4	71	88	2	4858	4614	**10491**	56,0	83,1
TOT	**1279**	**102**	**776**	**4913**	**8081**	**1086**	**8456**	**1127**	**15299**	**1641**	**5739**	**2103**	**646**	**2311**	**3509**	**987**	**790**	**6402**	**45847**	**111404**	**58,8**	**80,3**
Attractiveness	8,8	2,0	2,7	2,8	34,7	28,5	16,5	8,5	18,5	16,2	30,9	4,0	1,4	3,0	24,5	36,2	18,6	24,1	-	-	-	-

Table 2.9: Origin (University of bachelor's graduation within four years) - destination (University of master's enrollment within five years) matrix. Absolute values and rate of stayers. University students enrolled in Italy in 2008/09.

BA university	University of master's enrollment												
	UniBo	**Cattolica**	**UniTo**	**La Sapienza**	**UniPd**	**UniMi**	**PoliTo**	**Bicocca**	**Cà Foscari**	**UniMoRe**	**Other**	**%ST**	**Total**
North	480	417	302	74	279	360	93	303	253	167	1911	84,1	**29251**
Center	272	70	97	234	93	35	111	32	55	25	1391	80,2	**12222**
Chieti e Pescara	22	9	15	13	24	10	-	3	7	1	95	75,3	**805**
Federico II	23	6	6	26	7	2	23	5	2	1	172	87,6	**2199**
UniSalento	42	11	27	14	10	4	15	13	-	7	121	68,8	**845**
UniBa	31	20	12	21	21	16	2	7	3	-	95	78,4	**1056**
UniSalerno	12	1	2	22	3	3	16	1	-	3	130	75,5	**788**
UniCal	28	2	9	16	3	8	9	4	-	11	97	79,3	**905**
Other South	56	21	23	72	19	9	52	9	5	18	597	73,0	**3267**
South	214	70	94	184	87	52	117	42	17	41	1307	77,4	**9865**
UniMe	17	6	7	11	20	10	4	10	1	6	103	69,3	**635**
UniCa	16	3	30	7	14	6	15	2	3	3	91	68,6	**605**
UniCt	28	17	19	10	5	14	49	5	5	-	78	67,0	**696**
UniPa	22	13	22	14	15	5	45	7	9	11	150	65,5	**908**
Other Islands	14	7	28	5	11	7	0	6	1	-	99	57,5	**419**
Islands	97	46	106	47	65	42	113	30	19	20	521	66,1	**3263**
Total	**1063**	**603**	**599**	**539**	**524**	**489**	**434**	**407**	**344**	**253**	**5130**	**81,0**	**54601**

Table 2.10: Origin (University of bachelor's graduation within four years) - destination (University of master's enrollment within five years) matrix. Absolute values and rate of stayers. University students enrolled in Italy in 2014/15.

BA university	**University of master's enrollment**												
	UniBo	**Cattolica**	**UniTo**	**La Sapienza**	**UniPd**	**UniMi**	**PoliTo**	**Bicocca**	**Cà Foscari**	**UniMoRe**	**Altro**	**%ST**	**TOT**
North	758	772	530	132	482	564	173	429	422	287	3136	77,5	**34149**
Center	502	178	185	388	167	134	151	57	75	119	2070	72,5	**14640**
Chieti e Pescara	76	17	42	48	24	6	4	10	10	17	180	60,6	**1102**
Federico II	46	28	14	62	11	19	61	8	3	7	245	80,1	**2533**
UniSalento	38	10	43	27	9	17	7	17	2	13	159	60,1	**857**
UniBa	66	46	56	31	38	19	1	25	8	16	241	61,6	**1424**
UniSalerno	34	16	15	39	2	17	29	6	1	17	162	71,8	**1198**
UniCal	27	6	10	16	6	9	19	3	-	9	124	71,9	**815**
Other South	106	54	53	144	32	31	97	33	13	39	804	66,3	**4171**
South	393	177	233	367	122	118	218	102	37	118	1915	68,6	**12100**
UniMe	31	26	28	14	2	5	8	12	2	15	141	56,4	**651**
UniCa	52	13	32	16	26	8	29	6	6	14	119	53,8	**695**
UniCt	42	26	40	24	31	25	39	10	8	21	148	62,3	**1097**
UniPa	52	11	50	14	10	6	67	6	12	23	119	71,8	**1313**
Other Islands	18	15	22	14	15	7	10	9	4	4	131	58,6	**602**
Islands	195	91	172	82	84	51	153	43	32	77	658	62,4	**4358**
Total	**1848**	**1218**	**1120**	**969**	**855**	**867**	**695**	**631**	**566**	**601**	**7779**	**73,7**	**65247**

losing about 300 more students in 2014. Naples Federico II is the only university with a small number of incoming students, being a reference of the small university size of the Campania and southern area. Finally, the incoming mobility of Island universities is limited to exchanges between the same universities and some students coming from the neighboring Calabria.

Figure 2.3: Outflows (negative values) and inflows (positive values) of the top ten southern universities in the transition to the master's level by macro-region of master's enrollment. Outflows are movers to other universities; inflows are movers from other universities. Absolute values. University students enrolled in Italy in 2008/09 and 2014/15.

Finally, focusing on the incoming mobility of major destination universities in Figure 2.4, we see a substantial increase in outflows. The University of Bologna is confirmed as the major pole of attraction for master's enrollment, followed by Cattolica and the University of Turin. La Sapienza is a pole of attraction for graduates of southern and central universities. The Polytechnic University of Turin shows an almost equal distribution of the inflows based on the macro-region of origin, with a slight majority of students coming from the South. In contrast, the outflows from this university are not present.

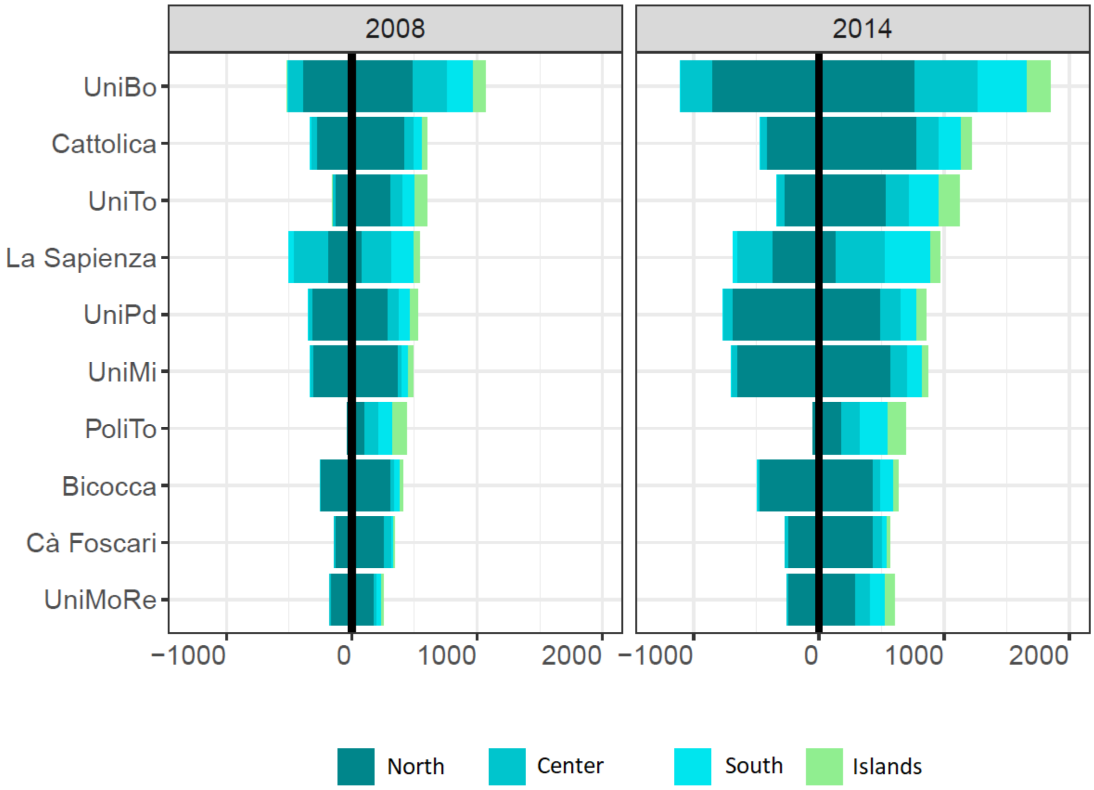

Figure 2.4: Outflows (negative values) and inflows (positive values) of the top ten central and northern universities in the transition to the master's level. Outflows are movers to other universities; inflows are movers from other universities. Absolute values. University students enrolled in Italy in 2008/09 and 2014/15.

2.5 Conclusions

This work describes Italian university mobility in the last decade. This mobility has only one direction: from the South to the Center-North. The ANS-U micro-data allowed us to analyze mobility from school to bachelor's enrolment and from bachelor's to master's. What emerges is an Italy split in two: the significant outflows of southern regions during the school-BA transition are supplemented by those in the transition to the master's level. Mobility from the South to the Center North increased over time in each transition. In fact, in 2017/18, fewer and fewer graduates decided to enroll at the master's level in the same region: one in three are leaving from the Islands and one in four from the South. As expected, no southern three-year graduate decides to return and enroll in a university in their area of origin. Moreover, those who graduate from a university in the South are more likely to continue their studies. At the same time, the continuation rate is lower in the North, probably due to the more receptive labor market. The preferences of students from the South, both at bachelor's and master's enrollment, seem to be increasingly directed toward the universities of Turin, Milan, and Bologna, denoting a strategic mobility choice that looks to the future access to the labor market.

A general reflection is unfortunately still valid after 50 years since the total opening of the university to students with any diploma. In fact, the Italian university system does not seem to be able to reduce the present gap between those who studied in a classical or scientific high school (which is known to be a proxy variable for socioeconomic status) and those who hold another degree, with this gap being more pronounced in the South. The school of origin represents a discriminating factor in all three aspects analyzed in this paper: the school-BA transition, the BA graduation rates, and the BA-MA transition.

The data presented here are, in the first instance, useful for universities in the South and Islands because they provide quantitative information on the mobility flows and destinations of those who decide to emigrate. They could serve to put in place, with analysis also at the sub-regional level, corrective policies, in concert with all other local and national levels, with the ultimate goal of reducing the flow of students leaving the South.

In conclusion, we note that mobility is one of the contributing variables, along with high grades in high school graduation and having attended a humanistic and scientific "liceo", to university success in terms of the time it takes to obtain a bachelor's degree and continuation to the master's level. Given the significant interaction among these variables, the results described should be taken with caution because it is not possible to isolate the effects of individual variables with the tools proposed in this paper; in fact, only regression models, combined with qualitative studies *ad hoc* could provide measures of the net effects of these components.

Chapter 3

An analysis of Italian university students' performance through segmented regression models: gender differences in STEM programs

This chapter is based on the work from Priulla, A., D'Angelo, N., Attanasio, M. (2021). *An analysis of Italian university students' performance through segmented regression models: gender differences in STEM courses.* Genus, 77(1), 1-20.

Abstract

This section focuses on studying gender differences in university performances in STEM programs in Italy. The aim is to investigate the relationship between the number of university credits earned during the first year (a good predictor of the regularity of the career) and the probability of obtaining a bachelor's within four years. To this aim, we used segmented regression models. Our analysis confirms that first-year performance is strongly correlated with bachelor's completion within four years. Furthermore, our findings show that gender differences vary among STEM programs, according to the care-oriented and technical-oriented dichotomy. Males outperform females in mathematics, physics, chem-

istry, and computer science, while females are slightly better than males in biology. In engineering, female performance seems to follow the male stream. Finally, accounting for other important covariates regarding students, we highlight the importance of high school background and students' demographic characteristics. The analysis concerns first-year students enrolled in 3-year STEM programs in Italian universities from 2008 to 2014. Data is provided by the Italian Ministry of University and Research (MIUR).

3.1 Introduction

In recent years, studies on student university experiences have been increasingly common (Salanova et al., 2010; Mega et al., 2014; Freeman et al., 2014). Several reasons have fostered this interest in Italy: first, the Bologna and the Bergen and Lisboa processes. Second, the reform of the Italian University system in 2001. Lastly, the central government's new funding system is based on the regularity of students' careers.

Major changes in the last decades have characterized the Italian higher education system. The reforms of the Italian University system developed along two main routes, leading to major changes in the binary single-tier structure of the Italian system. Firstly, a two-tier structure was introduced, in which students enroll in a first three-year cycle, namely the bachelor's degree, and, subsequently, may pursue a second two-year degree, namely the master's degree. Secondly, the reform considerably increased the number of fields students could choose.

The increase in horizontal stratification resulted in a strong variability of educational choices. In Italy, as in most other western countries (Mostafa, 2019), students are not particularly likely to enroll in STEM programs. Moreover, STEM programs have higher overall dropout rates than the other programs (Attanasio et al., 2018). Moreover, the poor female participation in STEM is still a relevant issue in Italy as in most worldwide education systems (De Vita and Giancola, 2017).

In this chapter, the aim is to investigate gender differences in university performance. As a measure of university performance, we consider the accumulation of university credits during the first year. As highlighted by Barone and Assirelli (2020), males and females tend to favor programs depending on their chances of succeeding in that program. Correll (2001) states that the perceived performance of individuals plays a crucial role in their inclination to pursue a career in a specific program. In addition, it frames the most important decisions in students' life cycle in their last years of high school when they consider university enrollment. However, in Italy, the most relevant decision in student careers can be identified in the choice of the degree program of enrollment. Therefore, our hypothesis is that gender differences can also vary among the different STEM programs: we are mindful of the recent work of Barone et al. (2019) that has claimed the presence of a care-technical divide within STEM programs.

The structure of the chapter is as follows. The literature review on gender differences in STEM is in Section 3.1.1, which is essentially an international glance with some references to the Italian context. The data is introduced in Section 3.2. Section 3.3 provides an exploratory data analysis. Section 3.4 outlines the modeling strategy based on the segmented regression models. The modeling results are reported in Section 3.5. Finally, in the conclusions, we try to connect our findings with the theory reported in Section 3.1.1.

3.1.1 Theoretical framework

In this section, we briefly describe some papers on gender differences in STEM, firstly concerning the high school and then the university context, with references to the international literature and the Italian experience. First, it should be noted that the definition of STEM can differ from country to country (Fan and Ritz, 2014). For example, definitions do not always include medicine, structural engineering, and sports science. Core STEM subjects typically include: mathematics; chemistry; computer science; biology; physics; architecture; and most engineering programs (UK-Parliament, 2020). In this work, we consider the abovementioned definition, excluding architecture.

Many authors analyze gender differences in STEM based on student performance in high school. There is extensive literature addressing the underperformance of males since the first schooling years. Females tend to do better than males in reading test scores, final grades, repetition at school, likelihood to choose academic tracks in high school, tertiary education attendance, and bachelor's graduation rates (Legewie and DiPrete, 2012).

From a sociological point of view, an interesting explanation, even if it concerns the US, is given by Correll (2001). The author states that gender differences in mathematics do not seem responsible for female and male choices to enroll in fields requiring a higher level of mathematical competence. She argues that cultural beliefs about gender and mathematics affect the choices of males and females toward educational paths leading to

STEM careers differently. Indeed, the author claims that some individuals believe that males are better at math, though females are less likely than males to hold stereotypical views about mathematics. The main conclusions of her work showed that, since males tend to overestimate their mathematical competence relative to females, males are also more likely to pursue activities that will lead to STEM careers. "Therefore, if a girl believes that males are better at math, she might view mathematical competence does not match her female gender identity, leading her to doubt her mathematical ability and consequently to decrease her interest in careers requiring high levels of mathematical competence. However, it is only necessary that individuals perceive that others hold these gendered beliefs concerning mathematics to lead to biased self-assessments of their ability and reduce their performance" (Correll, 2001).

From a psychological point of view, it has often been hypothesized that females have an innate predisposition to prefer educational paths in the humanistic and caring disciplines. Many theories have been discussed over the years to explain these differences in educational choices. For instance, Sherman (1980) discussed how family, the school environment, and the teachers' attitudes all have a significant role in influencing how males and females develop distinct attitudes toward particular subjects and skills, which in turn affects their educational choices. An interesting explanation for low female interest in STEM comes from Barone et al. (2019). The authors highlight the absence of accurate high school information for students relative to the long-term job opportunities related to specific programs. In this regard, students frequently make decisions based only on their preferred fields of study or "dream jobs", frequently gender-stereotyped, without considering the financial benefits or potential career opportunities. Moreover, according *care-technical* divide. Some fields of study prepare students for care jobs, while others can address students to a care job like teaching as a second-best option, such as some scientific fields like mathematics and biology (Barone et al., 2019). According to this divide, females are not underrepresented in all STEM programs: on the one hand, they are less in the most technical ones, such as engineering and computer science; on the other hand, they exceed males in biology or healthcare-related programs, fields historically related to the traditional female stereotype. For this reason, we will consider STEM programs separately to examine the gender divide better. Considering all the STEM programs together would be misleading because biology is indeed more care-oriented than, for instance, computer science.

The stereotypical divide between male and female fields and occupations is mirrored in university outcomes. Although females usually perform better at university than males, they may face more severe difficulties in STEM, leading them to switch to a non-STEM program in the first years and, with regards to some science programs, to quit their university career altogether (Attanasio et al., 2018). A possible explanation for this phenomenon is given by Hall and Sandler (1982), which defined STEM programs, especially engineering ones, as a "chilly climate" for female students. The author states that science faculties express higher expectations for male students or make females feel their ambitions are less important than their male colleagues. It is worth noting that other theories exist, such as the rational choice theory, which argues that individuals tend to prefer educational options that enhance their chances of success (Becker, 2003; Barone and Assirelli, 2020). This theory conceptualizes gender differentiation as an outcome of socialization processes and rational choice factors (Gabay-Egozi et al., 2015). Additionally, this idea claims that students who are more focused on their careers are less likely to enroll in care-oriented programs. Fewer females choose a more technical job path due to females favoring soft sectors where career prospects are less important.

One of the main focuses of international literature is on student performance in math tests. This is because mathematics can be seen as a proxy for STEM ability and then for future university success. In Italy, several papers look at how males do better than females in math tests, and some explanations have been suggested to try to explain this gap. These studies are mainly based on INVALSI tests administered to students throughout the schooling years (Giofrè et al., 2020; Cascella et al., 2020). For instance, even after accounting for individual and family background traits, females consistently underperform males on maths tests in Italy, one of the nations with the largest gender gap. These results also show how the average gender gap increases with age and becomes larger among top-performing children. Therefore, females' underperformance in mathematics could explain the tendency for females to follow non-scientific careers (Contini et al., 2017).

At the university level, some papers have highlighted the importance of differentiating the analysis of student performance based on the field of study. For instance, Cheryan et al. (2009) investigates the determinants of participation in computer science programs, showing that the interest in these programs is influenced by exposure to environments associated with computer scientists. The conclusion drawn in their paper is that changing stereotypical computer science environments could inspire a new interest in pursuing this specific program choice. Eccles (2007) analyzes why females continue to be underrepresented in the physical sciences and engineering in universities. The author suggests that the main explanation for gender differences in the physical sciences and engineering occupations is the importance placed on different types of occupations by males and females. In biological sciences programs, classified by Barone et al. (2019) as care-oriented, little attention has been paid to the performance of females compared to males or perceptions of stereotype threat (Lauer et al., 2013). Biology programs are considered an exception among STEM fields since they are female-dominated. In particular, Simon (2010) studied gender differences in knowledge and attitude towards biotechnology. His studies follow those suggested by Correll (2001), in which more knowledge in biotechnology decreased students' probability of being pessimistic about science. However, more knowledge in biotechnology led to a greater probability of pessimism for females. Eddy (2014) states that: "Often, gender differences are assumed to be present only in fields where males outnumber females and where there is a strong emphasis on math, but we are seeing it in undergraduate biology classrooms that do not focus on math - where females make up about 60% of the class - indicating that this could potentially be a much more systemic problem. Likely, this is not unique to physics or biology, but rather true of most undergraduate classrooms".

3.2 Data description

In this section, we use the ANS-U database. Here, each first-year student enrolled in a STEM degree program of an Italian university represents a statistical unit/record, which can be divided into two main parts: the first regarding high school background, and the second, divided into k parts (each one representing an academic year), which contains variables on their university career. This way, we can analyze student performance and its relationship with university completion.

Cohorts of students are analyzed in four-year time intervals, allowing for a follow-up looking at their progress from enrollment to potential bachelor's completion. We consider the 2014 cohort, which was, at the moment of writing, the most recent available cohort allowing us to observe bachelor's completion within four years.

Students enrolled at online universities are excluded from the study because of the different structures of non-online universities. We include students enrolled at both private and public universities. That distinction is not essential since STEM programs provided by private universities in Italy are limited and, therefore, not comparable to those of public universities. Additionally, we do not exclude dropout students from our research because doing so would result in overestimating the bachelor's completion rates. Moreover, high school grades in scientific subjects could help us understand university performance, but this information is not present in the ANS-U data. Previous high school variables and other personal characteristics are available, and they are named "admission covariates" due to their availability since the first university enrollment. Finally, as already said, a limitation of this work is the absence of information regarding family socioeconomic background in the ANS-U database.

3.3 Preliminary analysis

In this section, a brief preliminary analysis is conducted to investigate gender differences in university performance in STEM programs. It is divided into three parts: the first concerns university enrollment, the second the first-year performance, and the third the relationship between the type of high school attended and university performance.

First, we investigate the gender composition of STEM programs in Italy over the last decade. Table 3.1 shows the percentage of females enrolled in STEM courses in 2008 and 2014. Here, the 2008 cohort is included to get a temporal comparison. The variations between the two cohorts demonstrate that, regardless of gender, the total number of registered students has grown over the past six years. Nevertheless, this growth differs based on gender and field of study. The percentage of females fell by 2.3% even though the overall number of females was nearly unchanged. An increase is recorded in engineering (+2%), biotechnology (+1.3%), and natural sciences (+1.1%). The most striking decrease is in mathematics and statistics, where the percentage of female students decreased by more than 5%. A decrease is also observed in chemistry (-3.4%) and physics (-2.4%). Furthermore, disparities in gender composition might well be found for particular courses. Female students prefer to enroll in care-oriented programs, such as biology or biotechnology, with percentages of 72.4% and 65.3% in 2014. Conversely, the most male-dominated courses are also the most technical-oriented, namely computer science (13.6%) and engineering (24.5%).

Second, we examine student performance during their first year at university by examining the number of university credits earned. The median values of credits earned at the end of the first year for male and female students are shown separately in Figure 3.1, with

Table 3.1: Female students enrolled for the first time in STEM programs, and female enrollment rate of the 2008 and 2014 cohorts.

	Cohort					
	2008			**2014**		
Field of study	**F**	**%**	**M+F**	**F**	**%**	**M+F**
Biology	4785	*73,6*	**6505**	3945	*72,4*	**5446**
Biotechnology	2044	*64,0*	**3194**	1966	*65,3*	**3013**
Chemistry	1042	*48,2*	**2163**	1046	*44,8*	**2334**
Computer science	388	*14,7*	**2646**	544	*13,6*	**3992**
Engineering	5505	*22,5*	**24506**	6864	*24,5*	**28027**
Mathematics	1330	*61,8*	**2152**	838	*56,2*	**1490**
Natural sciences	1123	*48,5*	**2315**	1165	*49,6*	**2349**
Physics	610	*34,1*	**1787**	664	*31,7*	**2092**
Statistics	326	*46,6*	**700**	330	*41,5*	**796**
Total	**17153**	***37,3***	**45968**	**17362**	***35,0***	**49539**

the cohorts of first-year students enrolled in the academic years from 2008 to 2014. Student performance clearly differs based on the field of study and gender. On the one hand, as expected, female students perform better in the more care-oriented programs, such as biology and biotechnology. On the other hand, male students perform better in the most technical-oriented programs, such as physics and chemistry. Despite being considered one of the most "masculinized" programs, engineering does not show a significant gender gap in student performance, with female students showing slightly better results than their male colleagues. Furthermore, even after a slight improvement in recent years, natural and computer sciences are the programs where students exhibit the greatest difficulties. Nevertheless, there are no significant gender differences in performance in those programs.

Third, in Figure 3.2, we compute the BA completion rates by gender and type of high school attended. The BA rate is calculated as the number of students who graduated with a bachelor's within four years over the total enrolled students of the corresponding cohort. Some differences among programs can be observed related to the high school type. The separation between students from "licei" and others appears clear, especially compared to vocational school students. As expected, students from a scientific "liceo" seem to perform better in each scenario. Male students from technical schools achieve better results than those from a scientific "liceo" only in computer science programs. Students who completed their education abroad perform the worst, regardless of gender and program. Gender differences come up too. Females from humanistic "licei" perform better than their male counterparts, while male students from technical schools outperform females with the same

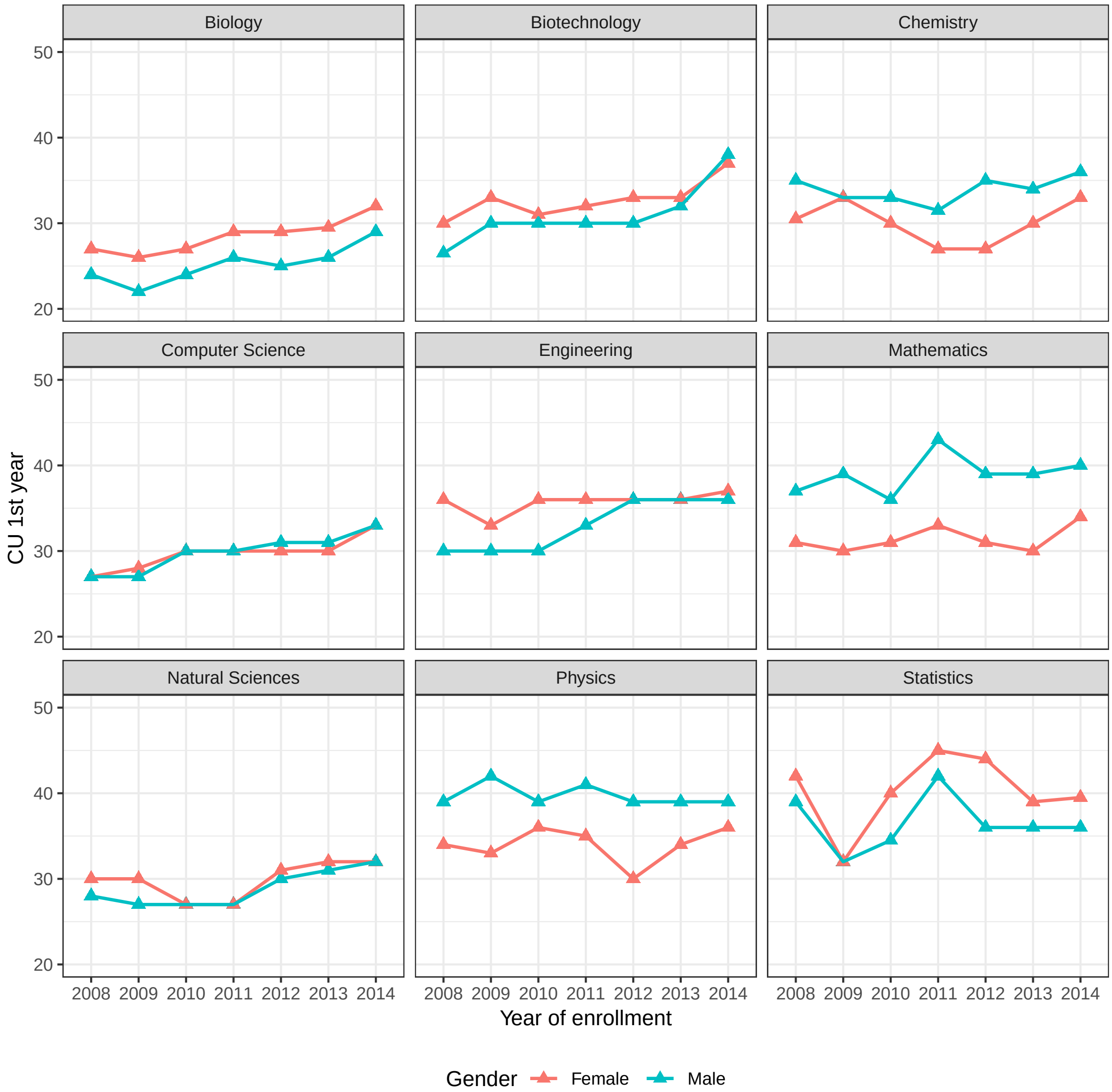

Figure 3.1: Median credits earned at the end of the first year (X-axis) by male and female students enrolled in STEM programs. Cohorts of first-year students enrolled from 2008 to 2014.

educational background. A possible explanation for this result can be addressed by the fact that females from humanistic "licei" are likely to be more involved than males from the same background, facing the challenge of enrolling in a STEM program.

3.4 Methods

This work aims to model the probability of graduation with a bachelor's within four years, conditioning on first-year university performance. In this respect, we consider a set of covariates:

- `CFU`: *first-year CFU*, which range from 0 to 60 (the annual credits in Italy);

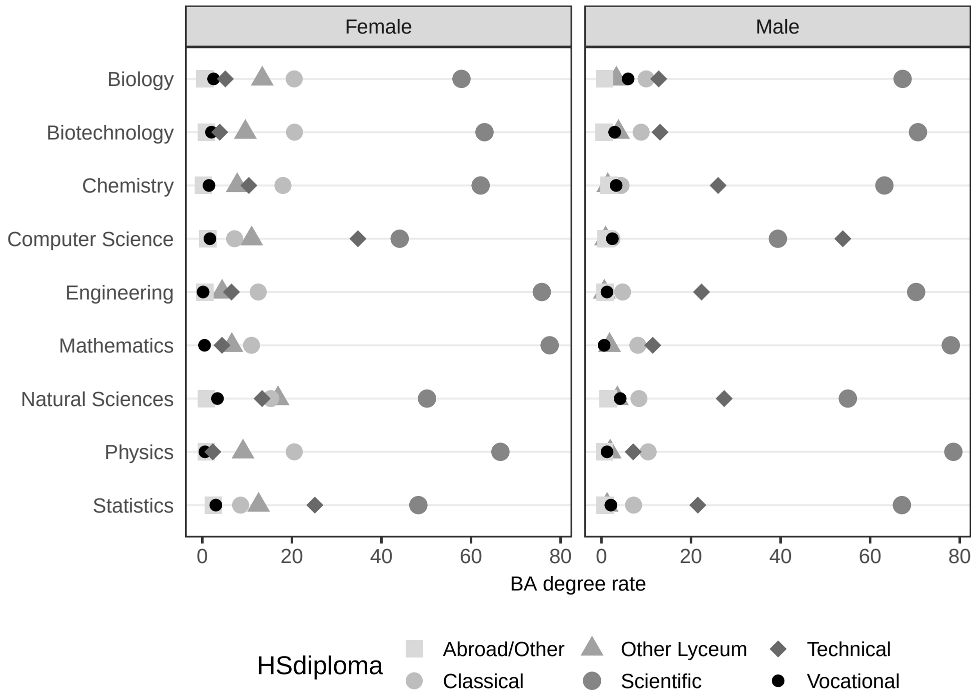

Figure 3.2: BA completion rates in STEM programs by gender and type of high school attended. Cohort of first-year students enrolled in 2014.

- `gender`;
- `age`: *age at enrollment*, dichotomized in ≤ 19 and > 19. We used this threshold since students with regular high school careers enroll at most at the age of 19;
- `macro-region`: *macro-region of enrollment*, categorized in North, Center, South, and Islands;
- `HSattended`: *type of high school attended*, categorized in humanistic and scientific "licei", other "licei", technical and vocational schools, and abroad/other;
- `HSfinalmark`: *high school final mark*, which ranges from 60 to 101, where 101 identifies "100 cum laude";
- `Program`: which identifies the 3-year STEM program of enrollment. Following the Barone et al. (2019) divide, programs are classified into two main groups:
  - *care-oriented* programs: biology, biotechnology, and mathematics;
  - *technical-oriented* programs: chemistry, computer science, engineering, natural sciences (which includes geology and environmental sciences), physics, and statistics.

Mathematics is included in the care-oriented group since most students enroll in this program aiming to take up a teaching career.

Previous works have suggested a strong relationship between first-year credits and the probability of bachelor's completion (Attanasio et al., 2013). However, little has been done to evaluate gender differences. In this section, we analyze the gender differences in student performance in STEM through the novel application of the *segmented regression models*. Recent papers deal with applications of segmented regression models in higher education (Geven et al., 2018; Li et al., 2019). However, to our knowledge, this work represents the first application of segmented regression models to predict university success.

All the analyses are performed using the `segmented` R package (Muggeo et al., 2008).

3.4.1 Segmented regression models

Segmented or broken-line models are regression models where the relationship between the response and one or more explanatory variables is piecewise linear and, as such, represented by two or more straight lines connected at unknown points. These models are a common tool in many fields, including epidemiology, occupational medicine, toxicology, and ecology, where it is usually of interest to assess threshold values after which the covariate effect changes (Ulm, 1991; Betts et al., 2007). Generally speaking, segmented regression models allow us to obtain a more synthetic representation and better interpretation, both analytically and graphically, of student university performance compared to other standard methods widely used in the literature. The main advantage of this approach is the straightforward interpretation given by two components: the *changepoint* (or the changepoints) and the slope (or the slopes). The changepoints are points in the range of the broken-line covariate, after which there is a change in the relationship with the response variable. Furthermore, those models represent a good trade-off between flexibility and computational burden, like the usual non-parametric approaches.

The segmented linear regression model can be expressed as

$$g(E[Y|x_i, z_i]) = \alpha + z_i^T\theta + \beta x_i + \sum_{k=1}^{K_0} \delta_k (x_i - \psi_k)_+ \tag{3.1}$$

where g is the link function, x_i is a broken-line covariate, and z_i is a covariate whose relationship with the response variable is not broken-line. We denote by K_0 the true number of changepoints and by ψ_k their K_0 locations in the range of x_i. The term

$(x_i - \psi_k)_+$ is defined as $I(x_j > \psi_k)$ that is $(x_i - \psi_k)I(x_i > \psi_k)$. The coefficient θ represents the non-broken-line effect of z_i, and β represents the effect for $x_i < \psi_1$, that is, the effect of x_i before the estimated changepoint. Finally, δ_k is the vector of the differences in the effects after the estimated changepoints. Throughout the paper, we only consider models with Gaussian iid errors $\epsilon_i \sim N(0, \sigma^2)$.

For estimation purposes, we consider a reparametrization of the segmented model. This has the advantage of a more efficient estimation approach via the algorithm discussed in Muggeo (2003) and Muggeo et al. (2008), fitting the generalized linear model iteratively:

$$g(E[Y|x_i]) = \beta_1 x_i + \sum_k \delta_k \tilde{U}_{ik} + \sum_k \gamma_k \tilde{V}_{ik}^-, \tag{3.2}$$

where $\tilde{U}_{ik} = (x_i - \tilde{\psi}_k)_+$, $\tilde{V}_{ik}^- = -I(x_i > \tilde{\psi}_k)$. The parameters β_1 and δ_k are the same as in Equation (3.1), while the γ are the working coefficients useful for the estimation procedure. At each step, the working model in Equation (3.2) is fitted, and new estimates of the changepoints are obtained via:

$$\hat{\psi}_k = \tilde{\psi}_k + \frac{\hat{\gamma}_k}{\hat{\delta}_k}$$

iterating the process up to convergence.

Let us consider, as an example, a model with a single changepoint ψ_1:

$$E\left[Y|x_i\right] = \beta_0 + \beta_1 x_i + \delta_1 \left(x_i - \psi_1\right)_+$$

This specification is particularly appealing, as it allows us to graphically represent the segmented relationship between the response and the broken-line covariates.

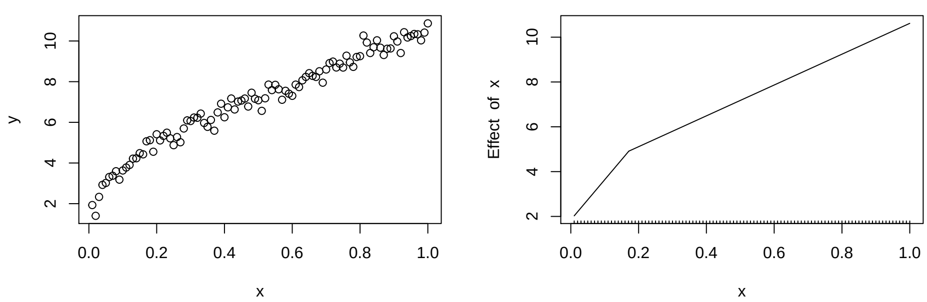

Figure 3.3: Simulated segmented relationship with a single changepoint.

In Figure 3.3, a simulated segmented relationship in the presence of a single covariate is shown. The left panel displays a non-linear relation between a covariate that takes $n = 100$ equispaced values ranging from 0 to 1, and the response variable given by

$$y_i = 2 + 15x_i - 8(x_i - 0.2) + \epsilon_i, \qquad \epsilon_i \sim \mathcal{N}(0, 0.3).$$

The right panel of Figure 3.3 displays the graphical output of the segmented model. The model finds the unique changepoint at $\hat{\psi}_1 = 0.2036$. The other parameters are equal to $\hat{\beta}_0 = 2.014$, $\hat{\beta}_1 = 14.839$ and $\hat{\delta}_1 = -8.220$. Here, $\hat{\beta}_1$ represents the left slope, being the effect of x_i before the estimated changepoint $\hat{\psi}_1$, that is when $x_i < 0.2036$. To obtain the right slope, that is the effect of x_i when $x_i > 0.2036$, we have to sum $\hat{\beta}_1$ and $\hat{\delta}_1$, obtaining $\hat{\beta}_1 + \hat{\delta}_1 = 14.839 - 8.220 = 6.6189$, being, of course, lower than the left slope.

Nevertheless, observing more than one changepoint in a broken-line relationship is reasonable. In this respect, the fundamental statistical problem is determining the number of changepoints. In this respect, the parameters to be estimated are the number of changepoints K_0, their locations ψ_k, and the broken-line effects β and $\boldsymbol{\delta}$. Typically, we would need to select the significant changepoints by removing the spurious ones. Indeed, whether the generic $\hat{\psi}_k$ is not significant, the corresponding covariate V_k should be a noise variable, as it would be $\hat{\delta}_k \approx 0$. The optimal fitted model will have $\hat{K} \leq K_0$ changepoints selected by any criterion.

Literature has been concerned with the problem of determining the "best" subset of independent variables. Two major approaches have been proposed to solve this problem: information criteria and hypothesis testing (Hocking, 1976). In this work, we refer to D'Angelo and Priulla (2020) for a complete description of the problem of estimating the number of changepoints and the criteria adopted. The authors propose a modified version of the usual procedure for choosing the number of changepoints based on sequential hypothesis testing. Its "validity" has been assessed through simulations, proving that the proposal correctly identifies the true number of changepoints outperforming all the considered information-based criteria competitors in the binomial case. Therefore, the procedure reported below will be performed throughout this work.

3.4.2 Sequential hypothesis testing procedure for the choice of K_0

An approach to select the number of changepoints is proposed in Kim et al. (2000), relying on a sequential hypothesis testing procedure. It consists of performing different tests, starting from $\mathcal{H}_0 : K_0 = 0$ vs $\mathcal{H}_1 : K_0 = K_{max}$, where K_{max} is fixed a priori. If $\mathcal{H}_0$ is rejected, the procedure tests for the next hypothesis by increasing by one the number of changepoints specified in $\mathcal{H}_0$ or by decreasing the one postulated under $\mathcal{H}_1$. D'Angelo and Priulla (2020) propose a different procedure to identify the correct number of changepoints. The procedure is based on sequential hypothesis testing using the pseudo-score and Davies' tests. The procedure starts from testing $\mathcal{H}_0 : K_0 = 0$, i.e. the model with no changepoints, vs $\mathcal{H}_1 : K_0 = 1$, i.e. the model with one changepoint. Depending on the tests' results, the procedure ends testing at most $\mathcal{H}_0 : K_0 = K_{max} - 1$ vs $\mathcal{H}_1 : K_0 = K_{max}$, and selecting up to K_{max} changepoints. Furthermore, we control for the over-rejection of the null hypotheses at the overall level α, using the Bonferroni correction α/K_{max} for the p-value. Of course, setting the Bonferroni correction to α/K_{max} is a conservative choice.

As compared to the procedure in Kim et al. (2000), the proposal of D'Angelo and Priulla (2020) has the advantage of not being limited to test for a maximum of additional *a priori* fixed changepoints. Moreover, the proposal of Kim et al. (2000) makes testing for more than two additional changepoints with the pseudo-score unfeasible because the current implementation of the pseudo-score test does not allow for testing for $\mathcal{H}_0 : K_0 = K$ vs $\mathcal{H}_1 : K_0 = K + 3$. D'Angelo and Priulla (2020) overcome this problem by accommodating any additional changepoints through the sequential procedure outlined below.

Steps of the procedure

Setting $K_{max} = 2$, the procedure is as follows:

1. Fit a segmented model to the data, with $\hat{K} = 1$ and test

$$\begin{cases} \mathcal{H}_0 : & \delta_1 = 0 \quad (K_0 = 0) \\ \mathcal{H}_1 : & \delta_1 \neq 0 \quad (K_0 > 1) \end{cases}$$

   using the Score or Davies' test. If $\mathcal{H}_0$ is not rejected, then the procedure stops estimating $\hat{K} = 0$. Otherwise, go to the next step.

2. Fit a segmented model with $\hat{K} = 2$ and test

$$\begin{cases} \mathcal{H}_0: & \delta_2 = 0 \quad (K_0 = 1) \\ \mathcal{H}_1: & \delta_2 \neq 0 \quad (K_0 > 2) \end{cases}$$

If $\mathcal{H}_0$ is not rejected, then the procedure stops at $\hat{K} = 1$, otherwise it stops at $\hat{K} = 2$.

In practice, the iterative procedure with Davies' test always stops as it gets $\hat{K} = 2$, even if the actual number can be larger. This is because Davies' test tests for at least an additional changepoint at each step. In our application, we define π as the probability of obtaining a bachelor's within four years. The coefficients α and λ represent the intercept and the slope of `CFU`, respectively.

The procedure to fit (3.1) is then outlined. We first fit the model (3.3):

$$\log\left(\frac{\pi}{1-\pi}\right) = \alpha + \lambda \texttt{CFU}_i \tag{3.3}$$

accounting for only the covariate `CFU`, to assess its effect on the probability of success.

Secondly, we fit a segmented logistic regression model to investigate whether some thresholds exist in the credits. This threshold would indicate a point after which a significant change in the probability of success is recorded. We include gender as the first and only non-segmented variable in the segmented model because we want first to assess the significance of this variable. To better analyze gender differences, we accommodate two instrumental covariates into the equation: $\texttt{CFU}_{male}$ and $\texttt{CFU}_{female}$. Then, the model has the following form:

$$\begin{aligned} \log\left(\frac{\pi}{1-\pi}\right) =& \alpha + \lambda_1 \texttt{CFU}_{male,i} + \lambda_2 \texttt{CFU}_{female,i} + \theta_1 \texttt{gender}_i \\ &+ \sum_{j=1}^{J} (\beta_j \texttt{CFU}_{j,i} + \sum_{k=1}^{K_j} \delta_{j,k} (\texttt{CFU}_{j,i} - \psi_{j,k})_+) \end{aligned} \tag{3.4}$$

where z_i is just the variable `gender`, and x_i are `CFU` for males and females.

The baseline profile is: { 0 for $\texttt{CFU}_{male}$ and $\texttt{CFU}_{female}$ } and {`female` for `gender`}. `Gender` is indexed by j, corresponding to two different segmented relationships, and K_{male} and K_{female} are the changepoints to be estimated for males and females. From now on, we will call this model the *marginal* model.

The segmented regression estimation procedure works plugging in $\hat{K}_j = 1, 2$ for $j =$

{$\texttt{male}, \texttt{female}$}, separately. In this way, we compare five models. Four out of five models are given by combining the two $\hat{K}_j$, and the fifth is the null model with no changepoints. Then, we apply the sequential hypothesis testing procedure outlined in Section 3.4.2 to select the “best” number of changepoints. In Figure 3.4, the broken-line relationship between the logit of the probability of success and credits for the baseline profile is displayed. The sequential procedure selects $\hat{K}_{male} = 1$ and $\hat{K}_{female} = 2$. The first changepoints are not distant; the two lines are roughly parallel after them, indicating no relevant difference in the relationship by gender.

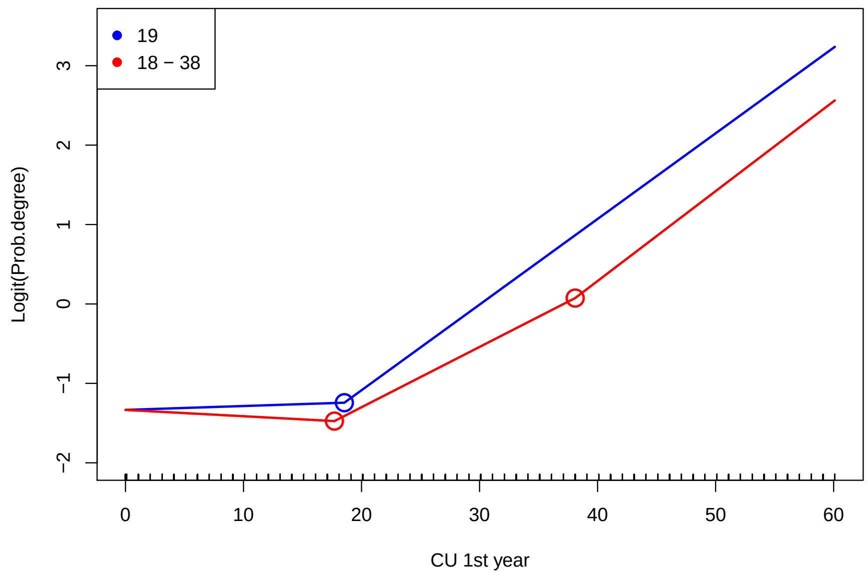

Figure 3.4: Segmented relationships between the logit of the probability of success (Y-axis) and the credits earned at the end of the first year (X-axis) of the marginal model in Equation 3.4. Males are represented by the blue broken-line and females by the red one. Cohort of first-year students enrolled in 2014.

The model fitted from Equation (3.4) can be further specified including the admission covariates. The new model is expressed in Equation (3.5).

$$\begin{aligned}\log\left(\frac{\pi}{1-\pi}\right) =& \alpha + \lambda_1\texttt{CFU}_{male,i} + \lambda_2\texttt{CFU}_{female,i} + \theta_1\texttt{gender}_i + \theta_2\texttt{macro-region}_i \\ &+ \theta_3\texttt{HSattended}_i + \theta_4\texttt{HSfinalmark}_i + \theta_5\texttt{age}_i + \theta_6\texttt{program}_i \\ &+ \theta_7\texttt{HSattended}_i * \texttt{gender}_i \\ &+ \sum_{j=1}^{J}(\beta_j\texttt{CFU}_{j,i} + \sum_{k=1}^{K_j}\delta_{j,k}(\texttt{CFU}_{j,i} - \psi_{j,k})_+).\end{aligned} \tag{3.5}$$

where z_i contains all the admission covariates, and x_i are CFU for males and females. We know there is a strong relationship between the covariate CFU and the admission covariates, as credits are determined at the end of the first year. However, the inclusion of z_i leads to a more than 20% improvement in the fitting due to the prolonged effect of z_i on the probability of bachelor's graduation. As happened for the marginal model, the sequential procedure selects $\hat{K}_{male} = 1$ and $\hat{K}_{female} = 2$.

Finally, since we are interested in analyzing how this relationship differs according to gender and the field of study, we estimate a stratified model. To avoid inserting several dummies, given by the couples $\{\texttt{program}_l * \texttt{CFU}_j \text{ with } l = 1, \ldots, 9 \; ; \; j = male, female\}$, we fit $L = 9$ program-specific segmented regression models, as in Equation (3.5). The program-specific model has the form:

$$\begin{aligned}\log\left(\frac{\pi_l}{1-\pi_l}\right) =& \alpha_l + \lambda_{1,l}\texttt{CFU}_{male,i} + \lambda_{2,l}\texttt{CFU}_{female,i} + \theta_{1,l}\texttt{gender}_i + \theta_{2,l}\texttt{macro-region}_i + \\ & + \theta_{3,l}\texttt{HSattended}_i + \theta_{4,l}\texttt{HSfinalmark}_i + \theta_{5,l}\texttt{age}_i + \\ & + \theta_{6,l}\texttt{HSattended}_i * \texttt{gender}_i + \sum_{j=1}^{J}(\beta_{jl}\texttt{CFU}_{j,i} + \sum_{k=1}^{K_j}\delta_{jl,k}(\texttt{CFU}_{j,i} - \psi_{jl,k})_+).\end{aligned} \tag{3.6}$$

Before proceeding to the results, it is important to stress that the estimated parameters of the chosen models cannot be considered in the usual "inferential" way since the data refers to a population. Nevertheless, the usual statistical procedures of model selection and estimation are used to understand the relationship among variables better.

3.5 Results

This section shows the results of both the marginal and the program-specific segmented regression models. The summary of the parameter estimates, the constant effects $\boldsymbol{\theta}$, and the broken-line effects $\boldsymbol{\psi}$ and $\boldsymbol{\delta}$ are reported in Tables 3.2 and 3.3.

First, the estimated parameters of the segmented model in Equation (3.5) are reported in Table 3.2. It can be noticed that the parameter estimates of CFU for both male and female students are very close. This indicates that the effect of credits does not differ by gender before the estimated changepoints. As for the admission covariates, results show a better female performance, attenuated by the interaction between gender and the type of

Table 3.2: Parameter estimates $\boldsymbol{\theta}$'s of the segmented regression model in Equation 3.5. Cohort of first-year students enrolled in 2014. Baselines are in brackets.

Variable	Estimate
Intercept	
	-1.96***
Gender (ref="Female")	
Male	-1.03***
CFU Male	
	-0.02*
CFU Female	
	-0.03*
HSattended (ref="Other liceo")	
Scientific "liceo"	-0.15*
Humanistic "liceo"	-0.17*
Technical institute	-0.39***
Vocational institute	-0.45***
Abroad/Other	-0.45*
Age at enrollment (ref="<=19")	
>19	-0.53***
Program (ref="Biology")	
Biotechnology	-0.26***
Chemistry	0.03
Computer science	-0.49***
Engineering	-0.29***
Mathematics	-0.32***
Natural sciences	0.08
Physics	-0.28***
Statistics	0.12
Macro-region of enrollment (ref="Islands")	
North	0.24***
Center	0.07
South	-0.18***
HS final mark	
	0.02***
HSattended * Gender	
Scientific "liceo"	0.28
Humanistic "liceo"	0.33*
Technical institute	0.39*
Vocational institute	0.39
Abroad/Other	0.50

high school attended. As expected, students from scientific and humanistic "licei" are more likely to graduate within four years than those with a vocational or technical background. Moreover, a higher HS final mark increases this probability. Unsurprisingly, students who enrolled late at university face the most difficulty completing a bachelor's within four years. Some differences can be highlighted in respect of the program: statistics students perform slightly better; students enrolled in computer science have substantial difficulties in completing the BA degree; natural sciences and chemistry students are close to biology ones in terms of performance. As for the macro-region, students enrolled at southern universities have an overall lower probability of success, followed by island students. Northern students perform the best.

Then, in Table 3.3, the other estimated parameters $\hat{\beta}_m$, $\hat{\beta}_f$ and $\hat{\delta}_{m,1}$, $\hat{\delta}_{f,1}$, $\hat{\delta}_{f,2}$ concern the segmented variable `CFU`. Male students show $\hat{K}_m = 1$, located at $\psi_{m,1} = 18.85$. For

Table 3.3: Parameter estimates of the $\boldsymbol{\psi}$'s and $\boldsymbol{\delta}$'s of the segmented regression model. Cohort of first-year students enrolled in 2014.

Variable	Parameter	Estimate	S.E.
CFU Male	$\psi_{m,1}$	18.85	0.72
CFU Female	$\psi_{f,1}$	15.14	1.66
	$\psi_{f,2}$	29.22	1.66
CFU Male	$\delta_{m,1}$	0.12	0.01
CFU Female	$\delta_{f,1}$	0.08	0.02
	$\delta_{f,2}$	0.05	0.01

female students, we have that $\hat{K}_f = 2$, which are located at $\psi_{f,1} = 15.14$ and $\psi_{f,2} = 29.22$. In practice, when students earn less than 20 credits, the probability of success does not change, regardless of gender. As shown in Figure 3.4, after 20 credits, the male line is always above the female one, with a slight difference till 30 credits. After 30 credits, the two lines run in parallel.

We now look for differences among programs by interpreting the results of models fitted as in Equation (3.6). Table 3.4 shows the number of changepoints selected by the procedures for the course-specific segmented regression models. The procedures found one changepoint for males in all the STEM programs and two for females in only five out of nine programs.

Table 3.4: Number of selected changepoints in males and females credits in the course-specific segmented regression models using Davies' and Score tests.

		Program								
Test	**Gender**	**Bio**	**Biot**	**Chem**	**Comp**	**Eng**	**Math**	**Nat**	**Phy**	**Stat**
Davies	**Male**	1	1	1	1	1	1	1	1	1
	Female	2	2	1	1	1	2	1	2	2
Score	**Male**	1	1	1	1	1	1	1	1	1
	Female	2	2	1	1	1	2	1	2	2

The estimated changepoints are displayed In Figure 3.5. Then, the parameter estimates of the fitted models are reported in Tables 3.5, 3.6, 3.7, and 3.8.

In Table 3.8, the parameter estimates of the $\boldsymbol{\psi}$'s and $\boldsymbol{\delta}$'s of the segmented regression model for each program are reported. Some differences can be underlined in the analyzed relationship. Only some programs display a significant effect before the first estimated changepoint, indicating a decrease in the probability of success before the estimated threshold. This occurs in computer science, mathematics, and physics, with a negative coefficient only for female students. Conversely, the effect of credits before the estimated changepoint is significant and positive in engineering, natural sciences, and biotechnology programs only for male students. One possible explanation could be the significant number of students leaving before completing those programs. This leads to a lower estimated probability of obtaining a bachelor's within four years when a slight increase occurs after a low number of credits. The first changepoints are almost always located between 10 and 25 credits, except in biotechnology and natural sciences, for both males and females. The first female changepoints come before male ones, but for engineering. This indicates that females need fewer credits to raise their probability of success in those programs. In biology, biotechnology, engineering, and chemistry, the relationships between the probability of success and the credits do not show significant gender differences. Other programs, such as computer science and mathematics, highlight significant gender differences in favor of male students.

Tables 3.5, 3.6, and 3.7 report the parameter estimates of the admission covariates for the program-specific segmented regression models. The gender parameter is between -1 and 0 in all programs but computer science, natural sciences, and statistics. The interaction effects must be considered while interpreting such estimates. In the first three programs, the main effects are compensated for by the interaction effects. The estimates of the `CFU Male` and `CFU Female` covariates range between -0.10 and $+0.10$, but those in mathematics, computer science, and statistics have higher negative values. All the parameters for the remaining admission covariates are mainly between -1 and $+1$. Some larger negative values are estimated for students from technical and vocational schools. Besides, the most striking difference in the interaction is observed in computer science, where the parameters are all positive, save for students who attended high school abroad. This means that being male from a traditional "liceo" or a technical or vocational school leads to a higher probability of success. Finally, it is important to note that students enrolled in northern universities perform better in almost every program, save statistics and biotechnology where students enrolled in the Islands perform better.

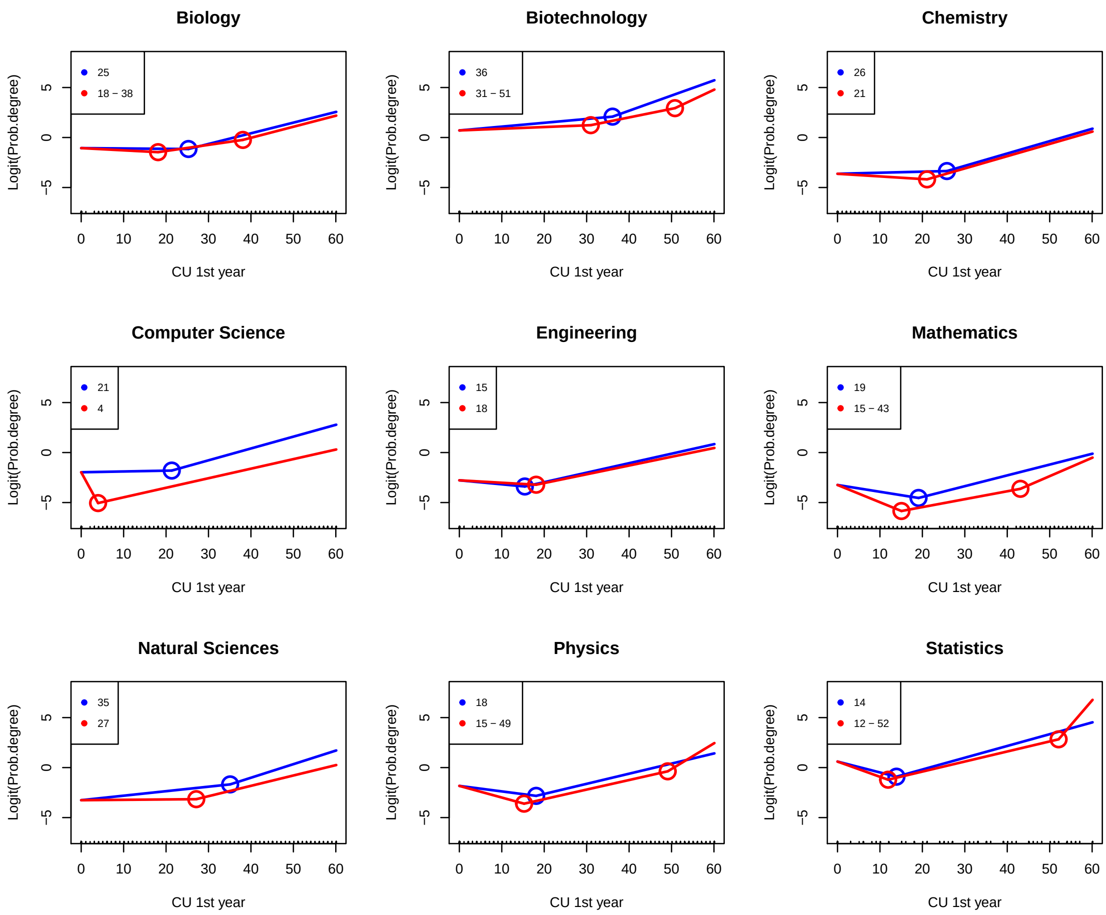

Figure 3.5: Segmented relationships between the logit of the probability of success (Y-axis) and the credits earned at the end of the first year (X-axis) of the program-specific segmented regression models in Equation 3.6. Males are represented by the blue broken-line and females by the red one. Cohort of first-year students enrolled in 2014.

3.6 Conclusions

STEM and gender have been a recent focus of worldwide academic literature, and quantitative studies on their relationship are essential for better understanding this topic. In this work, we restricted our analysis to the university and investigate gender differences in the probability of succeeding in STEM programs. In particular, we focus on the relationship between the bachelor's completion within four years and first-year performance.

The novelty of this work consists of a straightforward representation of the non-linear relationship between credits and the completion of the program through segmented models. Those models allow for identifying significant changepoints in credits accumulation during the student's first year at university, after which the probability of bachelor's completion increases. That relationship varies based on gender and the field of study at the university.

Table 3.5: Parameter estimates $\boldsymbol{\theta}$'s of the segmented regression models by university program. Baselines are in brackets. Cohort of first-year students enrolled in 2014 in biology, biotechnology, and chemistry. The asterisk indicates a corresponding p-value <0.05. Standard errors are in brackets.

	Program		
Variable	**Biology**	**Biotechnology**	**Chemistry**
Intercept	-1.06* (0.35)	0.71 (0.46)	-3.63* (0.62)
Gender (ref="Female")			
Male	-0.59 (0.45)	-0.85 (0.58)	0.00 (0.67)
CFU Male	0.00 (0.02)	0.04* (0.01)	0.01 (0.02)
CFU Female	-0.02 (0.02)	0.02 (0.01)	-0.03 (0.02)
HSattended (ref="Other liceo")			
Scientific "liceo"	-0.10 (0.11)	-0.68* (0.18)	-0.10 (0.29)
Humanistic "liceo"	-0.14 (0.13)	-0.56* (0.20)	-0.10 (0.32)
Technical institute	-0.35* (0.18)	-0.72* (0.30)	-0.36 (0.35)
Vocational institute	-0.51* (0.23)	-0.32 (0.39)	-0.85 (0.68)
Abroad/Other	-0.57 (0.46)	-0.29 (0.54)	-3.13* (1.19)
Age at enrollment (ref="<=19")			
>19	-0.41* (0.10)	-0.30 (0.16)	-0.74* (0.17)
Macro-region (ref="Islands")			
North	0.36* (0.11)	-0.23 (0.20)	0.17 (0.20)
Center	0.20 (0.11)	-0.17 (0.20)	0.09 (0.21)
South	-0.25* (0.11)	-0.07 (0.20)	-0.64* (0.24)
HS final mark	0.01 (0.00)	-0.01* (0.00)	0.03* (0.01)
HSattended * Gender			
Scientific "liceo"	-0.14 (0.36)	-0.24 (0.44)	-0.40 (0.63)
Humanistic "liceo"	-0.29 (0.41)	-0.45 (0.49)	-1.14 (0.72)
Technical institute	0.30 (0.42)	-0.20 (0.53)	-0.65 (0.67)
Vocational institute	0.45 (0.48)	-0.54 (0.70)	-0.21 (0.94)
Abroad/Other	0.51 (0.99)	-0.24 (1.18)	2.66 (1.44)

Our analysis confirms that first-year performance is strongly correlated to the achievement of a bachelor's within four years. This relationship often varies between males and females and is in line with Barone's divide between (female) care-oriented and (male) technical-oriented programs. This divide is consistent save in mathematics, where males outperform females. However, mathematics is included by Barone et al. (2019) in the (female) care-oriented group, probably because it was, in the past, a teaching-oriented program. Nowadays, a bachelor's in mathematics leads to a broader range of careers, with many technical and computer science jobs being taken up by math graduates. Therefore, we would suggest mathematics be considered both a care and technical program. Moreover, it is crucial to stress an upstream pattern in engineering, where female performance

Table 3.6: Parameter estimates $\boldsymbol{\theta}$'s of the segmented regression models by university program. Baselines are in brackets. Cohort of first-year students enrolled in 2014 in computer science, engineering, and mathematics. The asterisk indicates a corresponding p-value <0.05. Standard errors are in brackets.

	Program		
Variable	**Computer science**	**Engineering**	**Mathematics**
Intercept			
	-1.97 (1.05)	-2.78* (0.23)	-3.24* (0.79)
Gender (ref="Female")			
Male	-4.32* (1.09)	-0.75* (0.30)	-0.64 (0.95)
CFU Male			
	0.01 (0.02)	-0.04* (0.01)	-0.07 (0.05)
CFU Female			
	-0.78* (0.39)	-0.02 (0.01)	-0.17* (0.05)
HSattended (ref="Other liceo")			
Scientific "liceo"	-0.67 (0.39)	-0.01 (0.13)	0.36 (0.30)
Humanistic "liceo"	-0.64 (0.54)	-0.09 (0.14)	0.51 (0.37)
Technical institute	-1.13* (0.40)	-0.30 (0.16)	-0.06 (0.44)
Vocational institute	-1.36 (0.75)	-1.06* (0.53)	-0.61 (1.01)
Abroad/Other	0.89 (1.50)	-0.13 (0.36)	-24.93 (573.89)
Age at enrollment (ref="<=19")			
>19	-0.49* (0.11)	-0.58* (0.06)	-1.32* (0.33)
Macro-region (ref="Islands")			
North	1.40* (0.21)	0.10 (0.06)	0.55 (0.30)
Center	0.93* (0.23)	-0.07 (0.07)	0.56 (0.31)
South	0.34 (0.22)	-0.19* (0.07)	-0.02 (0.32)
HS final mark			
	0.02* (0.00)	0.02* (0.00)	0.04* (0.01)
HSattended * Gender			
Scientific "liceo"	1.75* (0.60)	0.33 (0.24)	-0.79 (0.84)
Humanistic "liceo"	2.17* (0.78)	0.42 (0.26)	0.17 (0.95)
Technical institute	1.92* (0.60)	0.53* (0.25)	-0.04 (0.93)
Vocational institute	1.90* (0.92)	1.18* (0.58)	0.46 (1.64)
Abroad/Other	-0.38 (1.63)	0.32 (0.45)	1.58 (847.23)

follows the male stream. There is a similar male-female performance in this important area, even if there are still few females.

Our findings show that gender differences vary greatly among STEM programs. These differences follow the care-oriented and technical-oriented dichotomy proposed in Barone et al. (2019), save, as noted, for mathematics. The probability of obtaining a bachelor's degree based on the number of credits earned at the end of the first year is higher for males in computer science and mathematics and slightly higher in natural sciences and biotechnology. The negative effect recorded for females could be related to their potential higher discouragement due to low academic achievement, and then to a higher probability of dropping out from those courses before completion. However, further investigation on the relationship between first-year performance and the choice to drop out from university

Table 3.7: Parameter estimates $\boldsymbol{\theta}$'s of the segmented regression models by university program. Baselines are in brackets. Cohort of first-year students in 2014 in natural sciences, physics, and statistics. The asterisk indicates a corresponding p-value <0.05. Standard errors are in brackets.

	Program		
Variable	**Natural sciences**	**Physics**	**Statistics**
Intercept	-3.26* (0.56)	-1.83* (0.69)	0.60 (1.42)
Gender (ref="Female")			
Male	-1.56* (0.55)	-0.59 (0.74)	-2.04 (1.63)
CFU Male	0.04* (0.01)	-0.06 (0.04)	-0.11 (0.14)
CFU Female	0.00 (0.01)	-0.12* (0.05)	-0.15 (0.16)
HSattended (ref="Other liceo")			
Scientific "liceo"	-0.02 (0.20)	0.07 (0.31)	-0.79 (0.60)
Humanistic "liceo"	0.06 (0.25)	-0.09 (0.35)	-0.55 (0.75)
Technical institute	-0.22 (0.25)	-1.24* (0.61)	-0.73 (0.61)
Vocational institute	-0.04 (0.37)	-0.62 (0.89)	-0.38 (1.24)
Abroad/Other	-0.64 (0.65)	0.07 (1.09)	2.11 (1.34)
Age at enrollment (ref="<=19")			
>19	-0.39* (0.13)	-0.68* (0.20)	-0.86* (0.27)
Macro-region (ref="Islands")			
North	0.40* (0.18)	0.64* (0.27)	-1.02 (0.65)
Center	0.19 (0.19)	0.23 (0.28)	-1.32 (0.69)
South	-0.13 (0.20)	0.01 (0.30)	-2.51* (0.73)
HS final mark	0.03* (0.01)	0.01* (0.01)	0.01 (0.01)
HSattended * Gender			
Scientific "liceo"	0.46 (0.44)	-0.32 (0.61)	1.83 (1.10)
Humanistic "liceo"	1.05* (0.53)	0.39 (0.67)	1.24 (1.30)
Technical institute	0.39 (0.47)	0.48 (0.82)	0.77 (1.12)
Vocational institute	0.15 (0.62)	0.50 (1.12)	0.91 (1.67)
Abroad/Other	0.90 (0.93)	0.16 (1.47)	23.02 (598.91)

is needed. Hall and Sandler (1982) suggested the interesting idea of the "chilly climate": the presence of university environments, such as some STEM courses, where females face more significant struggles in succeeding. It is interesting to stress that the theory of a "chilly climate" at university can be extended to high school in Italy. In this respect, Sherman (1980) states that school environment and teachers decisively shape the attitudes of males and females towards certain subjects and skills. This is confirmed by the gender composition of the scientific "liceo" and the technological-technical high schools in Italy, where females represent only, respectively, 43% and 17% of the graduates in 2019. These percentages show evidence that the gender gap in the scientific-technological fields is still present even though female participation has increased in the past 50 years.

Table 3.8: Parameter estimates of the $\boldsymbol{\psi}$'s and $\boldsymbol{\delta}$'s of the course-specific segmented regression models. Cohort of first-year students enrolled in 2014. Standard errors are in brackets.

Variable	Parameter	Biology	Biotechnology	Chemistry
CFU Male	$\psi_{m,1}$	25.27 (2.32)	36.10 (2.53)	25.82 (3.12)
CFU Female	$\psi_{f,1}$	18.12 (2.36)	30.97 (3.06)	21.10 (2.57)
	$\psi_{f,2}$	38.10 (3.08)	50.81 (2.52)	-
CFU Male	$\delta_{m,1}$	0.11* (0.02)	0.11* (0.02)	0.11* (0.02)
CFU Female	$\delta_{f,1}$	0.08* (0.02)	0.07* (0.02)	0.15* (0.03)
	$\delta_{f,2}$	0.05* (0.02)	0.11* (0.05)	-
		Computer science	**Engineering**	**Mathematics**
CFU Male	$\psi_{m,1}$	21.35 (2.70)	15.45 (0.85)	19.10 (3.99)
CFU Female	$\psi_{f,1}$	3.98 (1.24)	18.10 (1.36)	15.07 (2.23)
	$\psi_{f,2}$	-	-	43.10 (4.17)
CFU Male	$\delta_{m,1}$	0.11* (0.02)	0.13* (0.01)	0.18* (0.05)
CFU Female	$\delta_{f,1}$	0.87* (0.39)	0.11* (0.01)	0.25* (0.05)
	$\delta_{f,2}$	-	-	0.10* (0.05)
		Natural sciences	**Physics**	**Statistics**
CFU Male	$\psi_{m,1}$	35.09 (3.58)	18.10 (3.00)	13.89 (5.73)
CFU Female	$\psi_{f,1}$	27.12 (2.82)	15.25 (2.90)	11.92 (5.26)
	$\psi_{f,2}$	-	49.10 (3.61)	52.10 (2.54)
CFU Male	$\delta_{m,1}$	0.09* (0.02)	0.16* (0.04)	0.23 (0.14)
CFU Female	$\delta_{f,1}$	0.10* (0.02)	0.21* (0.06)	0.25 (0.16)
	$\delta_{f,2}$	-	0.16 (0.10)	0.39 (0.31)

Chapter 4

Gender differences in university students' trajectories in Italy

This chapter is based on the work from Priulla, A., Attanasio, M. (2023). *Gender differences in university students' trajectories in Italy.* Submitted.

Abstract

In this section, the aim is to investigate gender differences in university student trajectories in Italy, starting from the first enrollment to the bachelor's completion and the subsequent master's degree enrollment. It is known that females are better in terms of dropout rates, bachelor's completion rates, and final grades. However, little has been done to analyze the retention up to the master's level due to the lack of longitudinal data. We will use discrete-time multi-state Markov models to shed light on the factors most affecting students' decisions during their university careers. The analysis shows that factors related to the high school career and the field of study affect the university careers of males and females differently. The data concerns first-year students enrolled in 3-year programs in Italian universities from 2008 to 2020.

4.1 Introduction

Female advancement in educational attainment has been extraordinary over the last decades. New educational entries have exponentially risen after the Second World War, and now they exceed males in secondary school and academic achievement in most European countries (Salmieri and Giancola, 2020).

On average, across all the OECD countries, males made up only 45% of first-time entrants into tertiary education in 2019, with this share varying from less than 40% to 55%. Although more females than males are now tertiary graduates, the gap narrows in advanced levels of tertiary education, such as master's or doctoral programs. Around 18% of females are expected to enroll in a master's degree before the age of 30, compared to 12% of males. This gap further reduces at the doctoral level, where the average entry rate is around 0.9% for both males and females (OECD, 2021).

Nevertheless, vertical and horizontal gender segregation are still significant issues in higher education, especially in STEM degrees. Gender segregation can be described on two levels: the concept of *vertical* gender segregation is used to address female under-representation at higher levels of education, such as master's or doctoral levels; the concept of *horizontal* gender segregation is instead used to address the different educational choices of males and females.

The low female participation in STEM has been widely discussed in the worldwide academic literature over the last decades (Charles and Bradley, 2002; Macarie and Moldovan, 2015). Although the number of female STEM graduates increased in the EU after the expansion of higher education in the mid-2000s, horizontal segregation in STEM seems to be an unsolvable issue (Caprile et al., 2015). On the one hand, males still strongly dominate the fields of Information and Communication Technologies (ICTs) and Engineering, with a share of around 70% and 61%, respectively, across OECD countries. In detail, Italy has one of the lowest shares of female new entrants in tertiary education enrolling in ICTs, less than 1% in 2019. On the other hand, the gender imbalance is inverted in health and education, where females are largely over-represented. Females represented more than 75% of new entrants in the field of education across OECD countries. Natural sciences (including biology, geology, physics, etc.), mathematics, and statistics are the only STEM fields where females represented more than 50% of new entrants. Vertical gender segregation can also be observed in those programs where female participation is higher. Despite the already noticed gender balance in natural sciences, mathematics, and statistics, where females made up around 54% of the bachelor's and master's graduates, they made up only 46% of doctoral graduates in 2019 (OECD, 2021). moreover, data from the 2020 Graduate Outcomes survey, conducted annually by the Higher Education Statistics Agency (HESA) on UK graduates 15 months after they completed their studies, shows that the share of males engaging in further education after bachelor's completion is higher than females. This regards especially ICTs and Business and Administration programs, where the difference is around 16 percentage points.

Research on gender differences in Italy has been flourishing due to the significant changes in the educational system over the last decades to boost gender equality and reduce segregation in educational choices (Colombo and Salmieri, 2020). Borgna and Struffolino (2017) studied the gender differences in early school leaving, and their findings show that males' greater likelihood to drop out before completion is positively associated with better opportunities for them in the labor market.

The recent literature in tertiary education has mainly focused on horizontal gender segregation regarding career choices (Triventi, 2010; Barone, 2011). In contrast, vertical segregation has been studied mainly in the labor market framework (Checchi and Peragine, 2010).

In this work, we aim to contribute to the existing literature analyzing gender segregation in the transition from the BA to MA level at university. This transition is essential because it represents a crucial turning point requiring students to make new adjustments to a radically new educational context (Romito et al., 2020; Vettori et al., 2021). In addition, in a context in which increasing importance is devoted to the gender gap in a labor market where, nowadays, more and more technical skills are required, understanding student transitions to higher levels of education could provide some valuable hints to explain the root of the problem and reduce this gap (Gerber and Schaefer, 2004).

However, little has been done to analyze this transition due to the lack of longitudinal data covering a student's entire university career. One of the few exceptions is the work from Enea (2016), that analyzed the regional mobility of Italian students at the MA enrollment concerning some demographic variables registered in the ANS database, such as gender, the type of high school attended, the macro-region of enrollment.

In this framework, the aim of this section is to investigate gender differences in Italian higher education following two directions: the first one is the well-known horizontal segregation in both STEM and non-STEM disciplines; the second one is the less-known vertical segregation in terms of propensity to enroll at the MA level in all the fields.

This chapter is structured as follows: in Section 4.1.1, a review of the recent academic literature on gender differences in STEM is presented; in Section 4.2, the data and variables are introduced; in Section 4.3 a preliminary analysis of Italian university student careers is carried out; in Section 4.4, the modeling strategy is explained. Discrete-time multi-state Markov models are used to analyze the whole university students' trajectory, focusing on three main events: dropout before completion, bachelor's completion, and master's enrolment; the results are shown in Section 4.5; finally, we try to link our findings with the recent literature in Section 4.6.

4.1.1 Theoretical framework

Research has widely addressed the problem of female participation in STEM, aiming at discovering the critical factors contributing to the consistent gender divide in terms of participation and performance in those programs.

The gender gap in education has been frequently linked to students' choices and learning orientations, to the influence of teachers, families, and peer groups which condition the educational careers of males and females since primary education (Crosnoe et al., 2008; Wang and Degol, 2013). The lower female participation at advanced levels of education is often referred to in the literature using a metaphor as a "leaky pipeline", which carries students from secondary school through university and up to the labor market. The concept of a "leaky pipeline" in life sciences was promoted in the United States in the 70s, and it is usually used to describe female under-representation at higher levels of education (Clark Blickenstaff, 2005; Miller and Wai, 2015). Leaks can occur at different stages of student university trajectories: students can either drop out before completion or conclude their studies without enrolling in a MA degree after BA graduation.

Dropout before university completion is a problem for STEM programs since many students withdraw or change their field of study at specific points of their careers (Sadler et al., 2012; Watkins and Mazur, 2013; Ulriksen et al., 2015), especially in the first years of a university career. Universities focus mainly on recruitment, but this may not be enough if students starting a STEM career decide to leave before completion (Corbett and Hill, 2015; Blackburn, 2017).

Weeden et al. (2020) highlight the importance of general social processes in STEM educational outcomes. In particular, the authors considered as possible explanations for gender differences in STEM outcomes: prior academic achievement, family-work orientation, self-assessed maths ability, and occupational plans. However, their findings show that only occupational programs are strongly influential since females are much less interested in a potential STEM career.

In the US, Maltese and Cooper (2017) investigated the experiences responsible for triggering and maintaining STEM interest. Based on a sample of 8000 students, the authors concluded that interest in STEM subjects is mainly generated during the years spent in high school. The authors also show that males' interest generates more independently, while females are more influenced by external factors that convince them to pursue and persist in STEM, such as peers' and teachers' support, as well as the grades obtained in scientific subjects. Similar results come from Gabay-Egozi et al. (2015). Using data from an original survey of curricular choices of Israeli high school students, the authors showed that young girls are less likely to enroll in STEM unless they are encouraged by their parents and peers and/or if they expect higher utility and success in those careers.

The gender divide in tertiary education is mainly reflected in the field of study choice. The work from Cheryan et al. (2017) has underlined the importance of not considering STEM programs as a whole block because treating those as a homogeneous field would be misleading. Looking at the differences among STEM disciplines may point out the more complex ways in which inequalities persist in the context of expanding educational opportunities. Females remain over-represented in certain fields, such as education and health, but under-represented in the more technical ones, such as ICTs or Engineering. The gender gap in career choices can be explained by the stereotypes deeply internalized in worldwide cultural beliefs (Cheryan, 2012). Those programs where the gender gap is more evident are those stereotypically considered the most masculinized (Ganley et al., 2018). Females who intend to pursue a STEM career are more likely to choose a field whose main interest is taking care of others, such as biology or natural sciences. On the contrary, males are more interested in technical subjects, such as Industrial Engineering or ICTs, where the gender gap has been further narrowing in recent years (Barone and Assirelli, 2020). The low proportion of females in STEM fields contributes to the spread of gender stereotypical threats towards scientific fields and the strengthening of gender gaps in career-related interests and choices (Wang and Degol, 2013).

For mathematics, things are different. Although educational literature often addressed mathematics as a male subject, as males generally outperform females in maths tests during schooling years (Contini et al., 2017), mathematics programs are almost gender balanced (De Vita and Giancola, 2017). Barone (2011) identifies mathematics as a care-oriented field because of the significant share of students aiming at taking up a teaching career. However, males remain strongly advantaged in career advancement in this field.

The job placement for mathematics graduates has radically changed in recent years due to the increasing demand for more computing and technical skills in the labor market (Miller and Hughes, 2017).

In Italy, university performance and retention are influenced by many factors related to the students' socio-demographic characteristics and previous schooling outcomes. High school background is among the key discriminating factors of university careers (Contini et al., 2018). The preparation provided by Italian high schools is strongly heterogeneous depending on the high school curriculum. Humanistic and scientific "licei" prepare students for a university career, while technical and vocational schools prepare students to enter the labor market. A different gender composition of the students also characterizes those schools. In Italy, females made up 43% and 34% of the high school graduates in 2018 in the scientific "liceo" and the technical schools, respectively. Those percentages reveal that the gender gap in the scientific-technological field is still present. Despite the lower female participation, the data on Italian high school graduates in 2018 show that 87% and 44% of females in the scientific "liceo" and the technical schools, respectively, decided to enroll at university, against the 83% and 34% of males with the same background.

Data on Italian academic professors show evidence of a persistent leaky pipeline in STEM. All the programs, also those with higher female participation, such as biology or chemistry, suffer from leaks in the pipeline. However, this phenomenon is less strong for the recent cohorts. For instance, the gender ratios (M/F) of full professors in the biological and chemical sciences decreased from 4.69 in 2000 to 1.92 in 2020, while the corresponding gender ratio in all the disciplines reduced from 5.82 in 2000 to 2.85 in 2020. The sharp reduction observed in the biological and chemical sciences programs is probably due to the large proportion of females entering those disciplines in recent years.

4.2 Data description

The data used for the analysis are the micro-level longitudinal data from the *Anagrafe Nazionale Studenti* (ANS). Throughout this section, we mainly focus on the students enrolled for the first time in an Italian university in the 2008/09 and 2014/15 academic years. Those are, respectively, the least and the most recent cohorts for which a follow-up of five years is available. Students are followed for five years aiming at studying university retention and success. This follow-up allows us to analyze the BA completion within the fourth year (as the MUR assumes four years as the regular time to complete a BA program), and five years for enrollment at the MA level.

We excluded first-year students enrolled in 3-year healthcare programs because of their negligible rate of MA-level enrollment.

4.3 Preliminary analysis

In this section, we carry out an exploratory analysis to investigate the gender differences occurring in three crucial moments of students' university careers: the transition from the first to the second year, the BA completion, and the MA enrollment. In detail, we analyze the differences occurring in those three career moments concerning the main characteristics of the students:

- *Type of high school attended*, a crucial factor in predicting students' performance and their university retention;
- *Field of study*, divided into 16 different fields (9 STEM, 7 non-STEM), partially following the ISCED classification (UNESCO, 2013). The criteria used for this classification differ slightly from the traditional ISCED classification since some classes have been separated according to the aim of this study. However, it is necessary since we know the heterogeneity in gender participation and performance in STEM and non-STEM programs (Cheryan et al., 2017; Priulla et al., 2021).

Throughout the exploratory analysis, the BA completion rates are calculated including all the dropouts that occurred over the years in the denominator. This is because, otherwise, the BA completion rate would be underestimated. The MA completion rates are not considered because the dropout rates at the MA level are almost negligible. As already said, throughout the analysis, we will consider the cohorts of first-year students enrolled in 2008/09 and 2014/15 in 3-year programs at Italian universities. The 2020/21 cohort, the last available cohort, is considered only in the initial analysis to have more recent insights on the gender composition of Italian universities.

4.3.1 The gender composition of the Italian universities

First, in Table 4.1, we show the percentages of females enrolled in STEM and non-STEM fields for three cohorts: 2008/09, 2014/15, and 2020/21. The 2020/21 cohort will no longer be considered throughout the study.

Although exceeding males in terms of enrollment in Italian higher education nowadays, females made up only around 38% of STEM new entrants in 2020. On the contrary, female participation is much higher in non-STEM programs, with percentages around 65% on average for the three cohorts under examination. As expected, there is substantial heterogeneity concerning the fields of study. On the one hand, females represent the majority in Biological and Life Sciences, with percentages around 69% for the three cohorts. On the other hand, females made up only 13.4%, and 23.7% of the students enrolled in ICTs and Industrial and Information Engineering in 2020, respectively. A turnaround is observed in Mathematics and Statistics, where males represent the majority after a reduction in female participation from 56.9% in 2008 to 46.8% in 2020. Regarding non-STEM programs, Social Services and Educational Sciences are the most female-dominated fields, with percentages higher than 90%. Business and Administration is the only gender-balanced non-STEM field, with a slightly higher percentage of male students recorded in each cohort.

Table 4.1: Percentage of females enrolled in STEM and non-STEM programs at Italian universities. University students enrolled in Italy in 2008/09, 2014/15, and 2020/21.

	2008		2014		2020	
Field of study	**% F**	**M+F**	**% F**	**M+F**	**% F**	**M+F**
Agriculture, forestry, and fishery	*29,4*	3387	*29,4*	3858	*26,4*	3541
Architecture and building	*50,2*	10081	*54,0*	6419	*56,0*	6707
Biological sciences	*68,3*	12821	*68,8*	10728	*69,2*	15800
Civil engineering	*29,4*	6646	*31,3*	4699	*34,3*	3839
Inform. and commun. technologies (ICTs)	*13,6*	4259	*13,4*	5427	*13,4*	7558
Industrial and information engineering	*19,0*	26010	*22,1*	30507	*23,7*	39730
Life sciences	*56,5*	4675	*56,8*	6640	*57,4*	7832
Mathematics and statistics	*56,9*	3963	*50,0*	3112	*46,8*	4764
Physical sciences	*39,6*	6351	*38,3*	7055	*42,2*	8526
STEM	***38,0***	**78193**	***36,9***	**78445**	***38,4***	**98297**
Arts	*73,3*	8860	*73,1*	6410	*72,7*	8425
Business and administration	*49,4*	41062	*47,3*	34934	*48,1*	42889
Humanities	*76,0*	29290	*74,2*	30497	*73,6*	36322
Journalism and information	*67,0*	7848	*67,4*	7683	*67,6*	11233
Social and behavioral science	*64,9*	23610	*62,2*	20670	*61,4*	25379
Social services	*90,1*	3267	*90,1*	2774	*90,7*	3945
Teacher training and education science	*92,4*	12297	*92,6*	8521	*92,8*	11589
Non-STEM	***66,5***	**126234**	***64,8***	**111489**	***65,1***	**139782**
Total	***55,6***	**204427**	***53,3***	**189934**	***54,0***	**238079**

Furthermore, it is useful to observe the gender-gap evolution at different levels of education in the time interval between 2008 and 2014. In Figure 4.1, we show the gender ratio (M/F) both at BA and MA enrollment in STEM and non-STEM programs. Here, the years on the x-axis represent the respective cohorts of enrollment. It is essential to remind that 3-year healthcare programs are excluded.

The gender ratio has recently increased on average, with evident differences according to the field of study. In non-STEM programs, the ratio is stable, with females being double the number of males at both career levels. In STEM programs, the ratios are always higher than one but differ at the BA and MA levels. After a steady decrease observed until 2012, the gender ratio at BA enrollment has increased for the last cohorts. Although the number of females enrolling in STEM increased in this period, the number of new male entrants was larger than the female ones. In contrast, the MA ratio has increased except for the most recent cohort. The decrease recorded for the last cohort in the MA ratio is encouraging. The growing trends in STEM fields at both levels, except for this decline, are disheartening for the future realization of gender equality, one of the primary goals of the 2030 Agenda for Sustainable Development (García-Peñalvo et al., 2022).

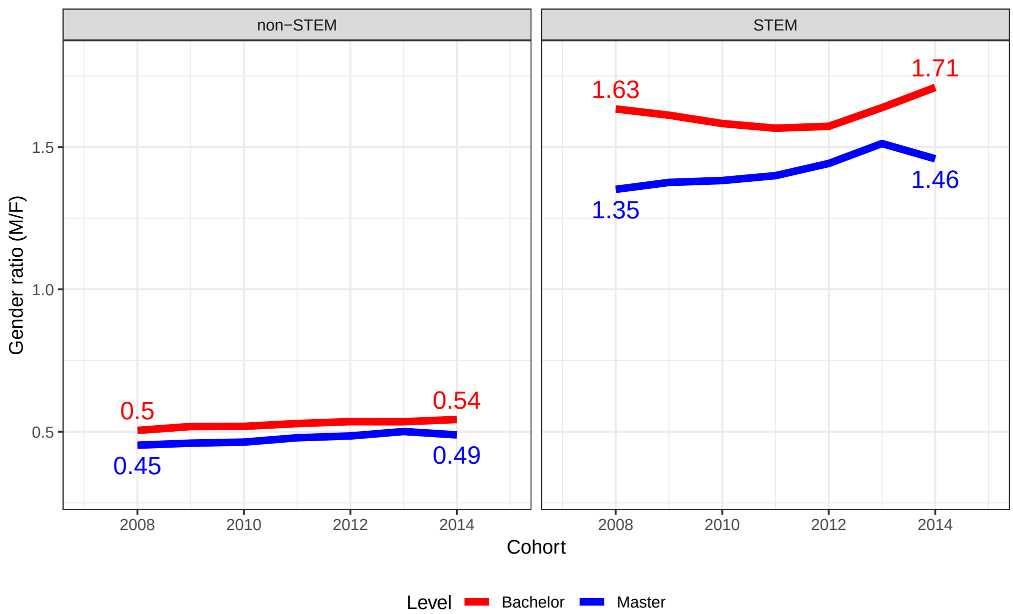

Figure 4.1: Gender ratio (M/F) in non-STEM and STEM fields at BA enrollment and MA enrollment. University students enrolled in Italy from 2008/09 to 2014/15.

4.3.2 The transition from the first to the second year

In this section, we aim to identify which students are at higher risk of early dropout from university. In this respect, we investigate the effect of the high school background, which is one of the factors most affecting student university careers (Attanasio et al., 2018).

In Table 4.2, we examine the transition from the first to the second year of a university career according to gender, high school background, and the field of study of first university enrollment. This step is crucial in student university careers since it represents the transition from the high school environment to the different and more "flexible" university environment. Students can be classified according to their choice in the first-second year transition:

- *Dropout*: a student no longer enrolled in an Italian university;
- *Mover*: a student who changed the program in the second year. A further classification is applied to identify students moving *Within* or *Outside* STEM or non-STEM fields;
- *Stayer*: a student who remains in the same field of study of enrollment.

It is important to specify that 5/6-year and 3-year healthcare programs are considered non-STEM programs.

Table 4.2: Dropout rates and change of field of study rates in the transition from the first to the second year, by gender, the field of study, and the type of high school attended. University students enrolled in Italy in 2014/15.

		Males					Females				
			Mover					Mover			
Field of study	HS attended	Dropout	Outside	Within	Stayer	Total	Dropout	Outside	Within	Stayer	Total
STEM	Scientific	*6,9*	*6,6*	*5,6*	*80,9*	**26318**	*4,8*	*9,0*	*7,4*	*78,9*	**15355**
	Humanistic	*7,2*	*12,6*	*6,1*	*74,1*	**2535**	*4,8*	*15,9*	*8,2*	*71,1*	**4450**
	Other "liceo"	*19,3*	*10,2*	*5,4*	*65,1*	**1154**	*14,5*	*12,9*	*9,8*	*62,8*	**3756**
	Technical	*22,9*	*3,6*	*4,2*	*69,2*	**16506**	*18,4*	*6,3*	*6,0*	*69,4*	**3909**
	Vocational	*35,2*	*3,8*	*3,8*	*57,2*	**2368**	*29,5*	*6,9*	*6,9*	*56,7*	**1086**
	Abroad/Other	*21,9*	*4,6*	*2,0*	*71,6*	**612**	*12,1*	*6,3*	*4,3*	*77,3*	**396**
	Total	***14,1***	***5,8***	***5,0***	***75,1***	**49493**	***8,9***	***10,1***	***7,6***	***73,5***	**28952**
non-STEM	Scientific	*10,8*	*3,0*	*4,2*	*82,0*	**15876**	*7,7*	*3,5*	*3,9*	*84,9*	**16467**
	Humanistic	*8,7*	*2,7*	*5,7*	*83,0*	**4441**	*6,0*	*3,1*	*4,5*	*86,4*	**11261**
	Other "liceo"	*18,7*	*1,8*	*5,2*	*74,4*	**3494**	*12,5*	*2,3*	*4,5*	*80,7*	**22487**
	Technical	*25,1*	*1,7*	*3,0*	*70,3*	**12269**	*18,7*	*1,4*	*3,4*	*76,5*	**15366**
	Vocational	*32,6*	*1,4*	*3,2*	*62,8*	**2361**	*27,0*	*0,9*	*4,0*	*68,1*	**5225**
	Abroad/Other	*24,5*	*1,4*	*2,8*	*71,4*	**789**	*19,1*	*1,4*	*3,3*	*76,2*	**1453**
	Total	***17,3***	***2,3***	***4,0***	***76,4***	**39230**	***12,9***	***2,4***	***4,0***	***80,7***	**72259**
Total		***15,5***	***4,3***	***4,6***	***75,7***	**88723**	***11,7***	***4,6***	***5,0***	***78,6***	**101211**

Results show that 13.5% of the students enrolled in Italian universities drop out after the first year. Those rates significantly differ by gender: 15.5% of males drop out, against 11.7% of females. Some differences can be observed concerning the field of study: STEM students show a lower average dropout rate and a higher propensity to change programs than non-STEM students. Focusing on STEM, it is interesting that the sum of the "half" failures (students moving to non-STEM programs) and the "thorough" failures (dropouts) are around 20% for both males and females. However, males and females show different behavior: females prefer to give themselves another chance moving to a non-STEM program (10.1%), while males prefer to give up (14.1%). As for the high school background, *fragile* students, namely those students facing more struggles throughout their university career, appear to be those from vocational or technical schools. In detail, students from vocational schools are more at risk of an early university dropout, regardless of the field of study. As expected, the results show that the selection process, which started in the transition from high school to university, proves that vocational schools are "unfit" for the university career. In contrast, as expected, students from scientific and humanistic "licei" show much lower dropout rates and a higher tendency to change from STEM to non-STEM, especially females from the humanistic "liceo".

4.3.3 University retention: who arrives at the end of the pipeline?

In this section, we analyze the BA completion rates and MA enrollment rates for the 2008/09 and the 2014/15 cohorts: the BA rate is calculated as

$$BA\% = \left(\frac{\text{no. of BA graduates within four years}}{\text{no. of BA enrollments}}\right) * 100; \tag{4.1}$$

the MA rate is calculated as

$$MA\% = \left(\frac{\text{no. of MA enrollments within five years}}{\text{no. of BA graduates within four years}}\right) * 100; \tag{4.2}$$

The rates are analyzed according to specific factors, such as gender, the type of high school attended, and the field of study.

The BA completion rates within four years for STEM and non-STEM students are reported in Table 4.3. Significant overall growth in the percentage of students graduating within four years is noteworthy, from 37.8% to 48.7% in 2014. Some differences come up concerning the field of study. STEM students show an overall lower BA rate than non-STEM ones, and this gap increases for the last cohort. In detail, the BA rates in STEM programs ranged between 29.1% in ICTs to 61% in Architecture and Building in 2014. In contrast, the rates in non-STEM programs ranged between 44.7% in Arts and 55.5% in Social and Behavioral Science. Yet, some gender differences can be observed: females show better BA completion rates even in most STEM programs. The only exception is Mathematics and Statistics, where the rates are almost equal. Female rates are always higher than 50% in non-STEM programs, while male ones are lower than 50%, except for

Business and Administration.

Table 4.3: BA degree rates within four years according to the field of study of enrollment and gender. The BA rate is calculated as the ratio between the total number of BA graduates within four years and the total number of BA enrollments. University students enrolled in Italy in 2008/09 and 2014/15.

	2008						2014					
	Males		Females		Total		Males		Females		Total	
Field of study	%	**Total**	%	**Total**	%	**Total**	%	**Total**	%	**Total**	%	**Total**
Agriculture, forestry, and fishery	*23,4*	2391	*29,1*	996	*25,1*	3387	*38,2*	2723	*45,0*	1135	*40,2*	3858
Architecture and building	*37,7*	5018	*54,0*	5063	*45,9*	10081	*52,3*	2951	*68,6*	3468	*61,1*	6419
Biological sciences	*24,6*	4070	*36,2*	8751	*32,5*	12821	*36,3*	3345	*46,7*	7383	*43,5*	10728
Civil engineering	*27,4*	4690	*37,2*	1956	*30,3*	6646	*32,5*	3229	*40,1*	1470	*34,9*	4699
Inform. and commun. technologies (ICTs)	*18,0*	3681	*25,8*	578	*19,1*	4259	*27,7*	4702	*38,2*	725	*29,1*	5427
Industrial and information engineering	*34,7*	21081	*48,1*	4929	*37,2*	26010	*40,1*	23767	*51,5*	6740	*42,6*	30507
Life sciences	*21,3*	2034	*28,1*	2641	*25,1*	4675	*34,3*	2870	*43,6*	3770	*39,6*	6640
Mathematics and statistics	*42,5*	1707	*42,2*	2256	*42,3*	3963	*46,1*	1556	*47,2*	1556	*46,7*	3112
Physical sciences	*38,5*	3833	*41,2*	2518	*39,6*	6351	*41,6*	4350	*46,1*	2705	*43,3*	7055
STEM	***31,6***	**48505**	***41,0***	**29688**	***35,2***	**78193**	***38,8***	**49493**	***49,4***	**28952**	***42,7***	**78445**
Arts	*25,8*	2369	*36,0*	6491	*33,3*	8860	*38,2*	1722	*47,1*	4688	*44,7*	6410
Business and administration	*35,4*	20779	*43,4*	20283	*39,4*	41062	*50,2*	18423	*58,1*	16511	*53,9*	34934
Humanities	*34,3*	7039	*43,5*	22251	*41,3*	29290	*42,2*	7867	*54,1*	22630	*51,0*	30497
Journalism and information	*34,3*	2588	*44,2*	5260	*40,9*	7848	*47,0*	2506	*58,5*	5177	*54,7*	7683
Social and behavioral science	*31,8*	8295	*45,0*	15315	*40,4*	23610	*47,6*	7808	*60,3*	12862	*55,5*	20670
Social services	*28,1*	324	*41,9*	2943	*40,5*	3267	*36,1*	274	*54,0*	2500	*52,2*	2774
Teacher training and education science	*23,7*	936	*37,9*	11361	*36,8*	12297	*39,2*	630	*55,8*	7891	*54,6*	8521
non-STEM	***33,6***	**42330**	***42,4***	**83904**	***39,4***	**126234**	***47,1***	**39230**	***56,2***	**72259**	***53,0***	**111489**
Total	***32,6***	**90835**	***42,0***	**113592**	***37,8***	**204427**	***42,4***	**88723**	***54,2***	**101211**	***48,7***	**189934**

In Table 4.4, we report the MA enrollment rates within five years for students who completed a BA degree within four years. We are aware that we are considering only the subset of the students completing the BA degree within four years. In this sense, this subset likely includes the "best" and "fastest" students. It is indeed known that the probability of enrolling at the MA level decreases when the time to complete the BA degree increases.

The overall MA rate remained stable at 67.1% in the two cohorts. Some differences can be noticed based on the field of study: STEM graduates show a higher propensity to enroll at the MA level than non-STEM ones, with rates around 77% and 62% in 2014, respectively. In detail, on the one hand, the MA enrollment rates of STEM graduates range between 52.9% in ICTs and 86.6% in Industrial Engineering. On the other hand, the rates of non-STEM graduates range between 41.1% in Journalism and Information and 70% in Social and Behavioral Science. As regards gender differences, there is an opposite pattern to the one observed at the BA level. Even if females are better in BA completion rates, males are more inclined to enroll at the MA level, showing a rate that is almost 7 points higher than females' one for the last cohort. Nevertheless, this is a misleading result

Table 4.4: MA enrollment rates by gender and BA completion field of study. The MA rate is calculated as the ratio between the number of students enrolled at the MA level within five years and the total number of BA graduates within four years. University students enrolled in Italy in 2008/09 and 2014/15.

	2008						2014					
	Males		Females		Total		Males		Females		Total	
Field of study	%	Total	%	Total	%	Total	%	Total	%	Total	%	Total
Agriculture, forestry, and fishery	60,6	545	60,5	261	60,6	806	63,8	1036	64,4	506	64,0	1542
Architecture and building	62,8	1786	64,9	2543	64,0	4329	61,0	1584	60,6	2424	60,8	4008
Biological sciences	79,1	770	84,9	2283	83,4	3053	80,3	1063	83,7	2792	82,8	3855
Civil engineering	87,0	1074	89,5	607	87,9	1681	83,8	914	83,3	515	83,6	1429
Inform. and commun. technologies (ICTs)	55,5	640	48,9	131	54,4	771	54,5	1289	43,7	229	52,9	1518
Industrial and information engineering	88,2	6088	90,1	1976	88,7	8064	85,9	8953	88,6	3171	86,6	12124
Life sciences	61,2	417	59,5	619	60,2	1036	60,3	923	60,0	1439	60,1	2362
Mathematics and statistics	80,9	640	84,3	791	82,8	1431	81,9	678	84,4	629	83,1	1307
Physical sciences	84,2	1374	81,8	870	83,3	2244	84,3	1722	80,1	1093	82,7	2815
STEM	**79,9**	**13334**	**78,2**	**10081**	**79,1**	**23415**	**78,2**	**18162**	**76,1**	**12798**	**77,3**	**30960**
Arts	48,3	606	51,5	2337	50,8	2943	56,7	682	58,1	2256	57,8	2938
Business and administration	70,6	7524	70,7	8855	70,7	16379	67,0	9496	66,2	9725	66,6	19221
Humanities	66,5	2446	62,9	9538	63,6	11984	65,5	3428	61,4	12097	62,3	15525
Journalism and information	44,8	951	43,9	2404	44,2	3355	39,1	1255	41,9	3110	41,1	4365
Social and behavioral science	69,8	2559	73,6	6786	72,6	9345	67,5	3756	71,5	7854	70,2	11610
Social services	41,6	89	42,5	1213	42,4	1302	51,0	96	52,8	1399	52,7	1495
Teacher training and education science	33,0	230	25,3	4354	25,7	4584	42,7	246	45,7	4513	45,5	4759
non-STEM	**66,3**	**14405**	**59,5**	**35487**	**61,5**	**49892**	**64,2**	**18959**	**60,8**	**40954**	**61,9**	**59913**
Total	**72,8**	**27739**	**63,7**	**45568**	**67,1**	**73307**	**71,1**	**37121**	**64,4**	**53752**	**67,1**	**90873**

since the gender gap reduces based on the field of study. As regards STEM programs, the main differences are recorded for ICTs and Physical Sciences graduates, where the rates are around 11 and 4 points higher for males, respectively. In detail, the gender gap in ICTs has widened from 2008 to 2014, following a decrease in the female MA enrollment rate. As for non-STEM graduates, the rate is always higher for females. The only exceptions are Humanities, where the rate is interestingly 3 points higher for males, and Business and Administration. This last result further supports our choice of differentiating the fields of study since the overall MA enrollment rate among non-STEM graduates was slightly higher for males.

Students' choice to engage in further education reflects their previous choices and performance. The international literature has indeed widely dealt with the transition from the first to the second year, both with small-scale surveys and theoretical models (Wintre and Bowers, 2007; Briggs et al., 2012), highlighting the importance of this transition in predicting student university success.

As already noticed, the Italian case is characterized by many factors influencing university success. Among those, the main factors are gender and the type of high school attended (Belloc et al., 2010; Contini et al., 2018; Priulla et al., 2021). Their influence seems to be still present in the transition to the MA level, as depicted in Figure 4.2. This figure shows the relationship between the dropout rates in the transition from the first to the second year and the rate of MA non-enrollment within five years, according to gender, the type of high school attended, and the field of study. At first sight, it seems that students already facing difficulties during the early stages of their university career struggle more in persisting up to advanced levels of education. Higher first-year dropout rates correspond to higher MA non-enrollment rates, especially in non-STEM programs. Students with a vocational background show the worst rates, while students from the scientific and humanistic "licei" show better rates. Those differences highlight the persistent gap in terms of university success among the different types of high school curricula. Some differences can be observed concerning the field of study. On the one hand, higher dropout rates are recorded for non-STEM students, whatever the high school background (17.3% and 12.9% for males and females, respectively). On the other hand, lower MA non-enrollment rates are recorded for STEM students (21.3% and 23.5% for males and females, respectively).

In Figure 4.3 we show the same relationship shown in Figure 4.2, conditioning on the university program.

The gender gap in MA-level enrollment is strongly heterogeneous among all the programs. Regarding STEM, females show higher dropout rates and a lower propensity than males to engage in further education, especially after obtaining a bachelor's in Computing and Physical Sciences. In addition, despite having higher dropout rates, females are more likely to enroll in a MA program after graduating in Mathematics and Statistics. Surprisingly, a slightly lower proportion of males choose to enroll at the MA level after obtaining a bachelor's in Industrial Engineering, with a rate of around 85%. More females persisting in such a male-dominated field is encouraging for reducing the gender gap in the labor market, but males still dramatically outnumber females. In contrast, the lower percentage of males enrolling at the MA level could be due to their higher possibility of finding a job without a MA degree in such a gender-unbalanced field.

As for Biological Sciences, a short digression is needed. Despite the limited dropout rate in the first-second year transition, Biological Sciences and some other scientific degrees suffer from a remarkable proportion of students changing to other programs, often belonging to the healthcare area. In Italy, access to healthcare programs is restricted

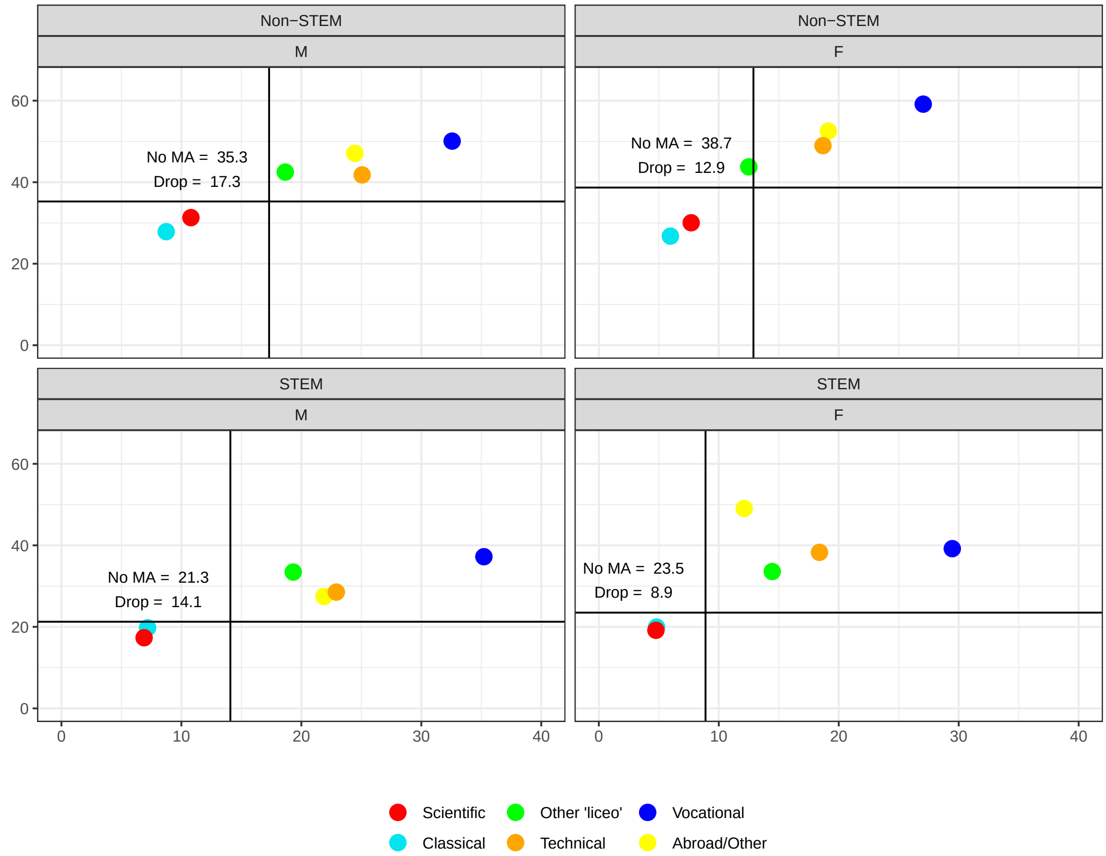

Figure 4.2: X-axis: dropout rates in the transition from the first to the second year; Y-axis: non-enrollment MA rates for BA in time graduates. The rates are calculated according to gender, the type of high school attended, and the field of study. The black straight lines represent the mean rates for each group. University students enrolled in Italy in 2014/15.

to a limited number of students based on a national-level admission test. Students who failed at the first attempt could decide to "park" in neighbor programs, such as Biological Sciences or Chemistry, aiming at reaching better preparation for the next-year test and at earning some credits that could be validated after a potential change to medicine in the following years.

In Figure 4.4, the same pattern recorded for STEM programs is observed for non-STEM ones. More than 15% of males leave earlier from each program, and they have lower rates of MA enrollment, except for Humanities and Business and Administration, the most numerous group. Differently, females' dropout rate in Teacher Training and Social Services, where female participation is overwhelming, is around 10 points lower than males. Those last results further highlight the primary importance of analyzing all the groups separately since considering only the overall rates of STEM and non-STEM fields would have been misleading.

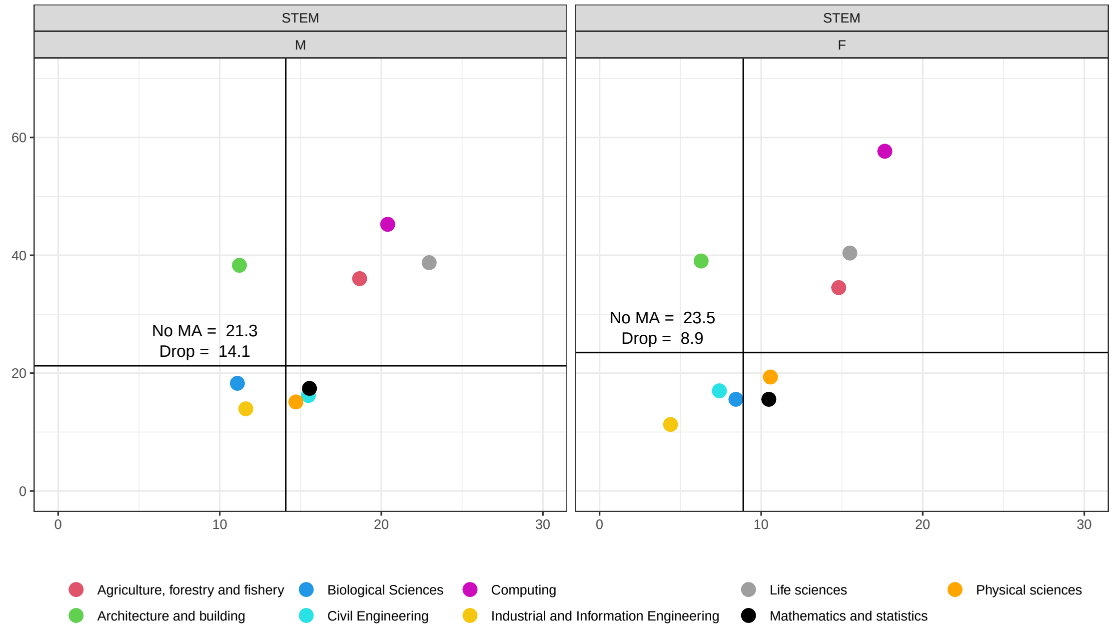

Figure 4.3: X-axis: dropout rates in the transition from the first to the second year; Y-axis: non-enrollment MA rates for BA graduates. By gender, STEM field of study, and type of high school attended. The black straight lines represent the mean rates for Xs and Ys in each group. University students enrolled in Italy in 2014/15.

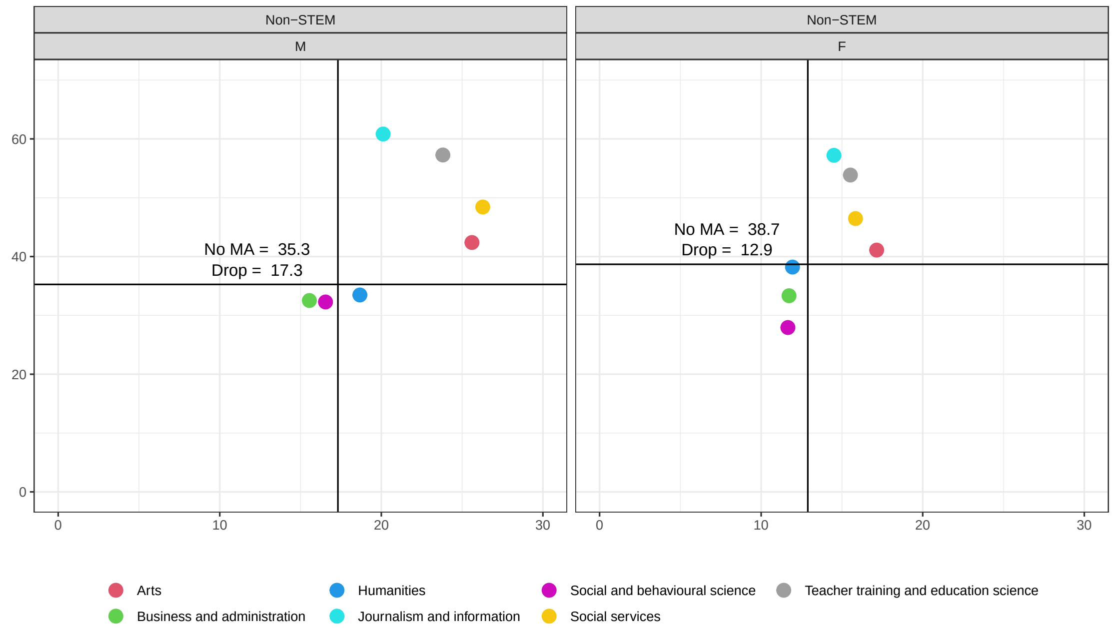

Figure 4.4: X-axis: dropout rates in the transition from the first to the second year; Y-axis: non-enrollment MA rates for BA graduates. By gender, the non-STEM field of study, and the type of high school attended. The black straight lines represent the mean rates for Xs and Ys in each group. University students enrolled in Italy in 2014/15.

In Table 4.5, we report BA degree and MA enrollment rates by gender, the field of study, and macro-region, for the 2014 cohort. The macro-regional gap in BA degree rates is evident, and this increases moving from southern to northern regions. Females always show higher BA rates than males, whatever the macro-region. This gap is larger in non-STEM programs, while it reduces in STEM programs, except for northern students. Nevertheless, the pattern just observed radically changes in terms of MA enrollment. Students from southern regions are more likely to enroll at the MA level, especially in non-STEM programs. In addition, the gender gap seems to be lower in southern regions, due to the lower MA enrollment rate of females who graduated in non-STEM programs. On the contrary, the gender gap in STEM programs is stable across the country. The Islands represent the only exception since the MA enrollment rate is higher for male students on average, but this pattern is reversed when conditioning for the field of study.

Table 4.5: BA degree and MA enrollment rates by gender, the field of study, and macro-region. University students enrolled in Italy in 2014/15.

Field of study	Macro-region	Bachelor's						Master's					
		Male		Female		Total		Male		Female		Total	
		%	n	%	n	%	n	%	n	%	n	%	n
STEM	North	*45,1*	7798	*57,2*	5545	*49,4*	13343	*76,9*	10961	*74,0*	7127	*75,8*	18088
	Center	*33,7*	2743	*44,4*	1990	*37,8*	4733	*80,3*	3383	*78,4*	2625	*79,5*	6008
	South	*30,8*	2009	*41,2*	1917	*35,0*	3926	*80,8*	2734	*78,7*	2294	*79,8*	5028
	Islands	*30,5*	784	*39,9*	629	*34,0*	1413	*78,4*	1084	*79,8*	752	*79,0*	1836
	Total	***38,8***	**13334**	***49,4***	**10081**	***42,7***	**23415**	***78,2***	**18162**	***76,1***	**12798**	***77,3***	**30960**
Non-STEM	North	*54,6*	7911	*63,6*	18391	*60,5*	26302	*59,8*	10381	*55,8*	21947	*57,1*	32328
	Center	*45,0*	3561	*53,6*	8006	*50,3*	11567	*69,2*	4506	*63,7*	8866	*65,5*	13372
	South	*37,2*	2245	*46,7*	6594	*43,5*	8839	*71,0*	3042	*69,8*	7280	*70,2*	10322
	Islands	*34,1*	688	*45,7*	2496	*42,0*	3184	*66,6*	1030	*67,1*	2861	*67,0*	3891
	Total	***47,1***	**14405**	***56,2***	**35487**	***53,0***	**49892**	***64,2***	**18959**	***60,8***	**40954**	***61,9***	**59913**
Total	North	*49,1*	15709	*61,8*	23936	*55,7*	39645	*68,6*	21342	*60,3*	29074	*63,8*	50416
	Center	*39,0*	6304	*50,9*	9996	*45,3*	16300	*74,0*	7889	*67,0*	11491	*69,9*	19380
	South	*33,8*	4254	*45,1*	8511	*40,0*	12765	*75,6*	5776	*71,9*	9574	*73,3*	15350
	Islands	*32,1*	1472	*44,2*	3125	*38,9*	4597	*72,7*	2114	*69,7*	3613	*70,8*	5727
	Total	***42,4***	**27739**	***54,2***	**45568**	***48,7***	**73307**	***71,1***	**37121**	***64,4***	**53752**	***67,1***	**90873**

4.4 The discrete-time multi-state Markov model

This paper analyzes student trajectories from first university enrollment to potential MA-level enrollment. It is known that, when dealing with longitudinal educational data, knowing the exact event occurrence time is fairly unrealistic. In the ANS database, each student is observed at the end of each academic year. For this reason, the times of event occurrences are measured as a discrete variable: the years from the first university enrollment up to the event occurrence. The starting point is the first university enrollment. Then, each student can experience three different events, as shown in Figure 4.5:

- *Ongoing*: a student still enrolled after five years;

- *Dropout*: a student who left the Italian university system before BA completion and did not re-enroll within two years, or a student who obtained at most 6 credits within the first two years (*unofficial dropout*);

- *BA degree*: a student who graduated as a bachelor in an Italian university within five years.

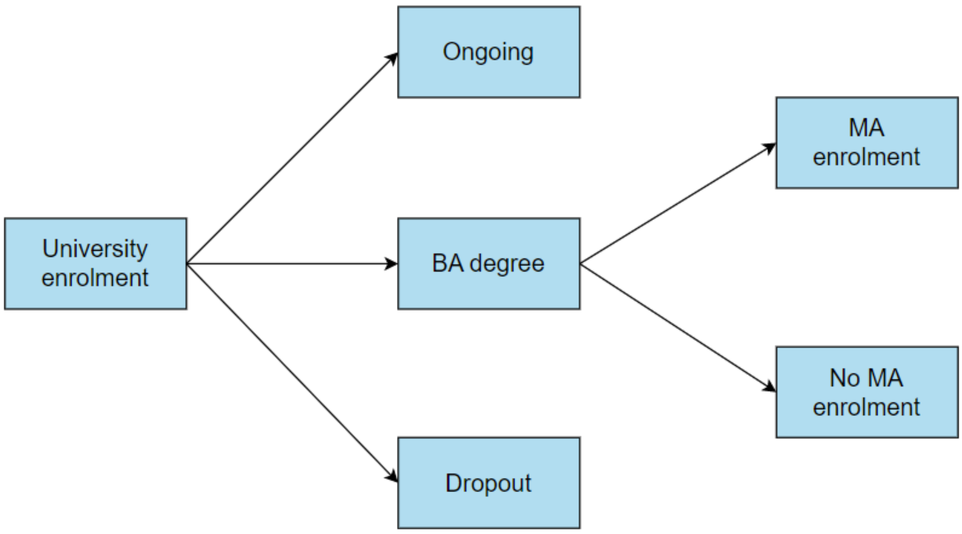

Figure 4.5: States and transitions of the multi-state Markov model.

Students who re-enrolled at university after dropping out are considered dropouts, meaning that *dropout* is an absorbing state. The BA graduation is an intermediate state after which students can enroll or not enroll at the MA level. Since the follow-up time is five years, students enrolling at the MA level after five years are considered not enrolled.

Differently from the model in Chapter 1, each event can occur at different times: dropout can occur during the entire follow-up time; the BA completion can occur from the third year onward; the MA enrollment can occur from the fourth year onward. However, as in the first chapter, our interest does not lie in the time of event occurrence but in the overall probability of moving from one state to another within a fixed time interval. In this respect, we define the following *transition probability matrix* **P**:

$$\mathbf{P} = \begin{array}{c} \begin{array}{cccccc} \text{Enrolled} & \text{Ongoing} & \text{Dropout} & \text{BA degree} & \text{MA enr.} & \text{No MA} \end{array} \\ \left[\begin{array}{cccccc} 0 & P_{12} & P_{13} & P_{14} & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & P_{45} & P_{46} \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right] \end{array} \begin{array}{l} \\ \text{Enrolled} \\ \text{Ongoing} \\ \text{Dropout} \\ \text{BA degree} \\ \text{MA enr.} \\ \text{No MA} \end{array} \tag{4.3}$$

The P_{ij} identifies the possible transitions. The rows of **P** satisfy the condition $\sum_{j=1}^{J} = 1$. We assume *time homogeneity* for the Markov chain. Hence, the transition probabilities do not depend on the time n.

The set of explanatory variables includes gender, the type of high school attended and final mark, the field of study of university enrollment, and the macro-region of enrollment. A dummy variable is also included to account for students who changed their field of study during their careers. Two-way gender interactions have been added with the macro-region, the high school attended, and the field of study. Moreover, a three-way interaction is included between gender, high school attended, and the field of study. As for the field of study and the macro-region, we account for changes during student careers. In other words, the field of study and the macro-region after BA completion might not match what was indicated at enrollment. Finally, to remove the bias due to the attraction of the healthcare area, we decided to exclude students who move to those programs at a certain point in their careers.

The model is applied only to the 2014 cohort, the most recent one with a five-year follow-up, since the completion times of the previous cohorts are generally longer, but the covariates' effects are similar.

4.5 Results

In this section, we report the results of the discrete-time Markov model. After comparing different models, we collapsed some fields of study to reduce the number of parameters: teacher training and education science and social services were collapsed into "education science and social services", business and administration and social and behavioral sciences into "economics and social sciences", and engineering programs into a single category.

The predicted transition probabilities are shown for three events: dropout in 4.6, BA completion in 4.7, and MA enrollment after BA completion in 4.8. Those probabilities are estimated by gender, the field of study of enrollment, and the type of high school attended. The other variables are fixed: a student with a final mark equal to 79, enrolled in a northern university, and who remained in the same degree program during his/her career. We consider this profile since it represents the "best" student. This means that, for example, the estimated probabilities of BA completion or MA enrollment would be similar for the other profiles but slightly lower. As for the final mark, which varies from 60 to 101, 79 represents the observed national average.

First, the estimated probabilities for the enrolled → dropout transition are in Figure 4.6. The probability of dropping out before completion is higher for male students and, overall, in STEM courses. Computing, mathematics and statistics, and physical sciences students are more likely to drop out before completion. In detail, mathematics and statistics is the only field where no significant difference between males and females is recorded. Surprisingly, engineering programs show the largest difference in favor of female students, followed by humanities and education science. As for the high school track, as expected, students from humanistic and scientific "licei" are less likely to drop out than technical students. Interestingly, the estimated probabilities differ based on the high school track. While the estimated probabilities are similar for humanistic and scientific licei, those for the technical schools are different: the gender gap observed for the traditional "licei" in mathematics and statistics and engineering is indeed reversed in technical schools, with females being more likely to drop out than males.

In Figure 4.7, we show the estimated transition probabilities for the enrolled → BA completion transition. Here, the pattern observed for the dropout transition is reversed: as expected, the probability of obtaining a bachelor's within five years is higher for female students and, overall, in non-STEM fields. The effect of the high school background is also noticeable. On average, students from technical schools have a lower probability of BA completion. Yet differences can be noticed according to the field of study and gender. The gender gap observed for students from humanistic "licei" enrolled in typical humanistic fields, such as arts and humanities or economics and social sciences, reduces for students from scientific and technical schools. Interestingly, males from technical schools are more likely than females to complete a bachelor's in engineering and mathematics and statistics fields.

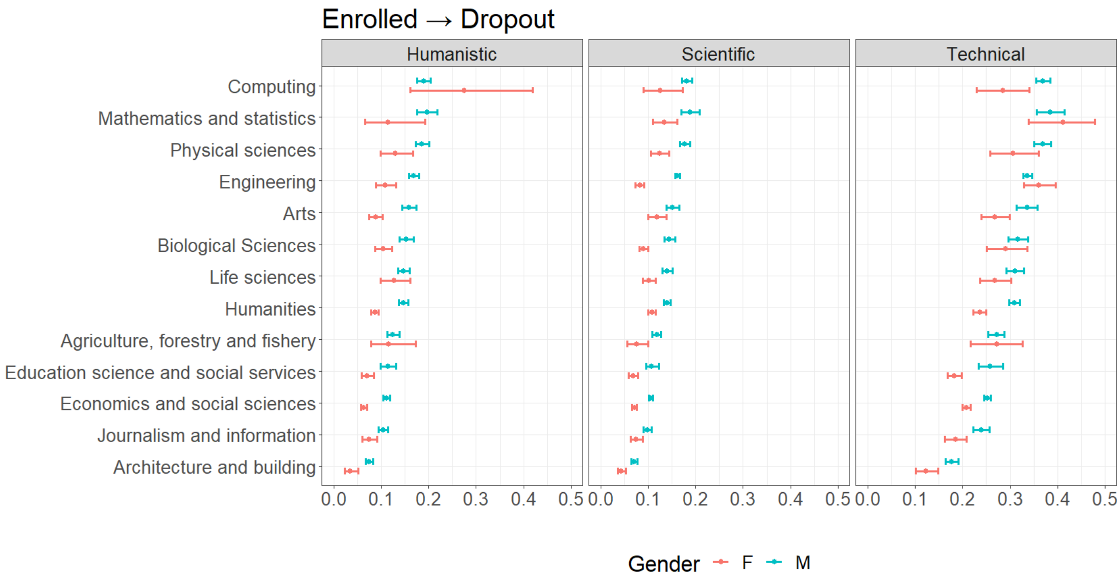

Figure 4.6: Estimated probabilities for the enrolled → dropout transition.

Finally, in Figure 4.8, we report the probabilities for the BA completion → MA enrollment transition. Male students seem more likely to enroll at the MA level. Moreover, students who graduated with a STEM bachelor are more likely to engage in further education. Only computing is different. This difference deserves some comments, in fact, this field of study has shown a sharp increase in labor market demand (National Academies of Sciences et al., 2018). Unfortunately, the available data do not allow us to verify what occurs after leaving the Italian university. Specific differences can be seen concerning the high school track attended. The parameters observed for traditional "licei" again differ from those for technical schools, where the gender gap in favor of male students is even more relevant. This regards especially humanities, economics, and mathematics and statistics fields.

4.6 Conclusions

Gender inequalities in higher education have been of interest in the international literature (Cheryan et al., 2017; Barone and Assirelli, 2020; Tandrayen-Ragoobur and Gokulsing, 2021; García-Peñalvo et al., 2022). Qualitative and quantitative studies have analyzed gender inequalities in retention at higher levels of education. However, little has been done to investigate gender differences in the transition to the MA level because of the unavailability of individual and longitudinal micro-data. Using the ANS longitudinal micro-data, we analyzed the gender differences in student trajectories in STEM and non-STEM programs

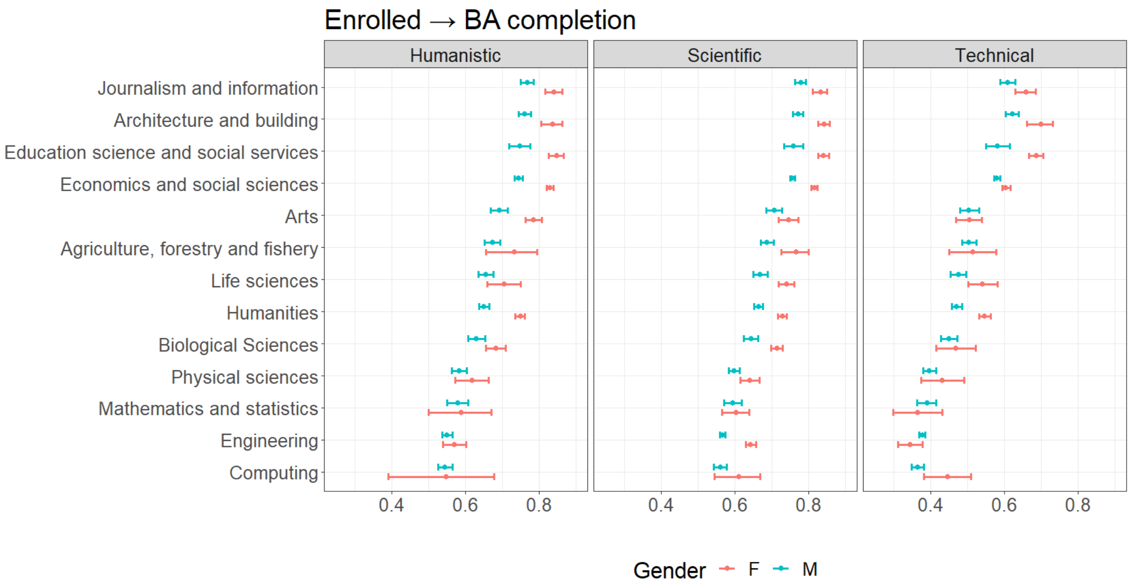

Figure 4.7: Estimated probabilities for the enrolled → BA completion transition.

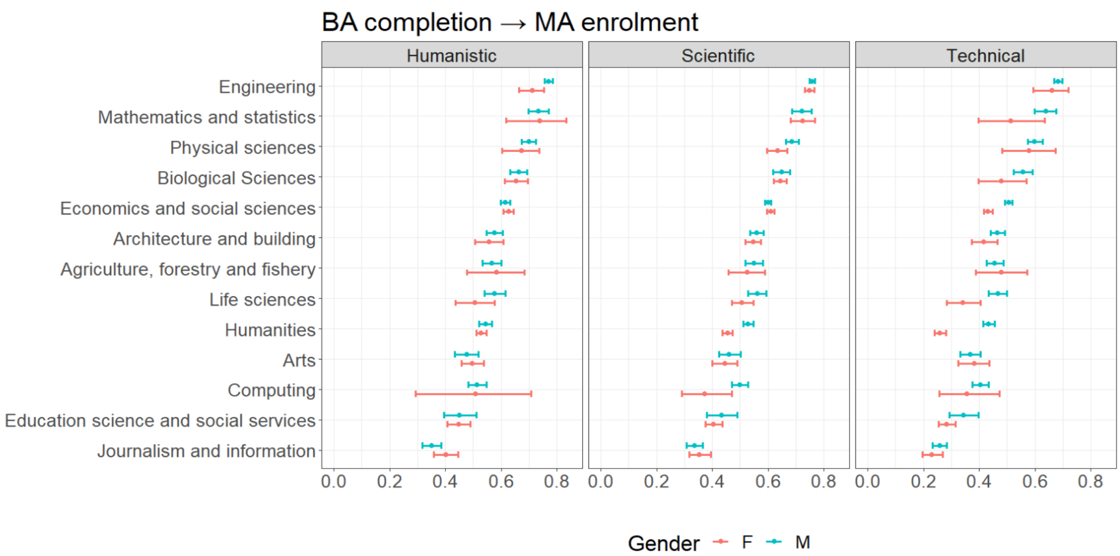

Figure 4.8: Estimated probabilities for the BA completion → MA enrollment transition.

from the first university enrollment to the MA level enrollment.

Our findings highlight some underlying interacting push factors that influence students' retention and performance at university. In this framework, the high school background is a "persistent" factor influencing the whole university career, from the first-second year transition to the MA enrollment. In general, it seems that, after 50 years since the bill allowing any diploma to access at university, the Italian higher education system has not narrowed the gap - in terms of university success - between students who attended scientific and humanistic "licei" and the others.

Although female participation in tertiary education has increased over the last decades, gender horizontal segregation is still relevant. In 2020, females represented only 38.4% of students enrolled in STEM programs. Moreover, our findings show that considering STEM programs as a whole block would be misleading: females comprised only 13.4% of first-year students in Computing, 23.7% in Industrial Engineering, 46.8% in Mathematics and Statistics, and 69.2% in Biological Sciences.

It is worth noting that there is a clear trend of vertical segregation of women in STEM fields, particularly when it comes to pursuing a master's degree. Despite a larger number of female students enrolling in master's programs, the percentage of women choosing to pursue STEM disciplines at this level is lower compared to their male counterparts. This phenomenon could be attributed to the prevalent horizontal segregation within the Italian high school system. In 2019, only 43% of graduates from scientific schools were female, while the numbers for industrial technical and vocational schools were 17% and 11%, respectively. The prevalence of male-dominated university environments could potentially discourage female students who come from a technical or vocational background from pursuing STEM fields (Hall and Sandler, 1982).

Using discrete-time Markov models, we analyzed student trajectories at the university. Our findings show that the probabilities of persisting at higher levels of education for males and females differ based on the field of study and the type of high school attended. Males are more likely to drop out and less to complete the BA degree within five years in all the courses. The only exception is mathematics and statistics, where no gender differences have been observed. Nonetheless, results show that this gender gap is reversed in some programs for students coming from technical schools. Further investigation on the eventual job destinations of the dropout students will be likely helpful in understanding the gender differences, as the job market is more male-oriented, especially for STEM jobs.

As for the transition to the MA level, we noticed that males, especially those from humanities and scientific "licei", have an overall higher probability of enrolling at the MA level, especially in Physical Sciences and, surprisingly, in Humanities. Interestingly, males are more likely to enroll in a MA program after obtaining a bachelor's in all the fields. The lower interest of females towards a MA enrollment in STEM could be linked to a possible discouragement of female students towards reaching leading positions due to the overwhelming number of males holding these positions in this field, but this needs more investigation.

Conclusions

Inequalities in higher education have been widely discussed in the international and Italian literature (Shavit and Blossfeld, 1993; Breen et al., 2009; Jackson, 2013; Contini and Scagni, 2013). In Italy, most works have used aggregate data or micro-data from single university archives (Clerici et al., 2015; Giudici et al., 2021). Other works have used survey data on students' universities careers (Belloc et al., 2010; Ballarino and Panichella, 2016; Aina et al., 2018; Aina and Casalone, 2020). The ANS-U database has allowed for addressing the problem of inequalities in higher education, with such detail rarely found in Italian literature. Thanks to the ANS-U data, we have achieved a deeper understanding of the transition to the master's level in terms of educational inequalities. Moreover, thanks to the recent linkage between ANS-U and INVALSI data, it was possible to deeply analyze the factors affecting the transition from high school to university. This large database allowed us to investigate educational inequalities from three perspectives: socioeconomic status, gender, and territorial differences.

In this framework, Chapter 1 deals with the effect of attending two scientific high school tracks, i.e., the traditional and applied science tracks, on two academic outcomes: the choice of enrolling at university and university first-year performance. The topic addresses interesting international research on mathematics's impact on academic success (Lee and Ready, 2009; Contini and Scagni, 2013; Wang, 2013; Poulsen, 2019). In Italy, research has mainly focused on the social stratification associated with school tracks and university performance. The novelty is twofold: the first regards the comparison between the traditional and applied science tracks, that absorb the highest percentage of high school students, around 25%; the second regards the application of a *discrete-time multi-state Markov* model to analyze educational biographies. To account for the hierarchical structure of the ANS-INVALSI database, we balanced the data for socioeconomic status, gender, and macroregional location of the schools, using a *multi-level propensity score* procedure.

The matching procedure enabled us to make a fair comparison between the two tracks by ensuring a balance in the composition of both groups regarding socioeconomic status, gender, and high school performance. However, the results indicate that this procedure reduces the number of students in the traditional track, particularly those with high performance levels. This leads to group alignment and, consequently, an underestimation of the students' academic performance in the traditional track, which is more significant for those in the South and those in STEM programs. In detail, according to the findings, students in the applied science track: i) are less likely to enroll in university and more likely to enroll in STEM programs than their traditional science peers; ii) perform similarly in STEM programs but worse in non-STEM programs than traditional science students. To the best of our knowledge, our findings highlight a distinction between these two high school tracks for the first time. Furthermore, significant macro-regional differences exist between the two tracks. In fact, the performance gap between students on the two tracks widens as one moves from north to south. This is most likely because the applied science track is still considered less prestigious in the South than the traditional one.

In Chapter 2, we provide a broad overview of student mobility in Italy from the South to the North, highlighting differences in academic performance and outcomes between movers and stayers. The South-North dualism represents an unending inequity affecting the Italian territory, mirrored in the higher education system in a vicious circle affecting social and economic development and strengthening student mobility. Because southern students do not return to their place of origin, unidirectional student flows from South to North are usually permanent. This phenomenon contributes to the South's human capital impoverishment, already plagued by high school dropout rates and low tertiary education enrollment (Ballarino et al., 2014). In detail, we used the ANS-U database to compare the university careers of movers and stayers in three steps: the first university enrolment, the bachelor's degree achievement, and the enrolment at the master's level. The main novelty lies in a spatial and temporal comparison between movers and stayers in terms of university performance. The findings highlight the positive impact of mobility on university achievement in terms of the time it takes to complete a bachelor's degree and the decision to pursue a master's degree. Because this chapter aimed to provide a descriptive analysis of student mobility flows in Italy, no statistical models were used.

Chapter 3 examined the gender differences in bachelor's degree achievement in STEM programs. The topic has been widely discussed in the international literature, based on the well-known humanistic-scientific divide. This divide has been reviewed by Barone (2011), who introduced the care-technical divide, in which females under-represented in most STEM fields, except for biological sciences. In this respect, we investigated potential gender differences in the single STEM courses. To this aim, we used *segmented regression models* (Muggeo et al., 2008) to analyze the non-linear relationship between first-year credits and program completion. The main feature of these models is the identification of significant changepoints in first-year credits accumulation, after which the likelihood of completing a bachelor's degree increases. Moreover, we proposed a sequential hypothesis testing procedure which allowed us to find the best number of estimated changepoints. The main findings show that the probability of obtaining a bachelor's degree based on the number of first-year credits is higher for males in computer science and mathematics and slightly higher in natural sciences and biotechnology.

Finally, Chapter 4 provided a more thorough analysis of gender differences in university trajectories in STEM and non-STEM courses. The main novelty of this work is the availability of longitudinal micro-data that allow the analysis of the entire university path from enrolment to the master's level. Studying this transition is essential because it represents a crucial turning point and the last step before entering the labor market. In detail, we investigated the gender differences regarding three events: dropout, bachelor's completion, and master's degree enrolment. Furthermore, literature has mainly addressed the problem of gender differences focusing on STEM programs (Cheryan et al., 2017; Griffith, 2010; Macarie and Moldovan, 2015). In this paper, we decided to consider non-STEM programs for a more detailed overview of the existing horizontal and vertical segregation in Italian tertiary education. As in Chapter 1, we used a *discrete-time multi-state Markov model* to define a complete transition probability structure from the first university enrolment to the master's enrolment. The results have shown that the pipeline in STEM, but also non-STEM programs, is affected by female leaks in the transition to the master's level. Females are less likely to enroll at the master's level in almost all fields. These results are in accordance with recent literature that show, despite not describing to the Italian framework, the underrepresentation of females at higher levels of education (Buchmann et al., 2008; Leemann, 2010; García-Peñalvo et al., 2022). Moreover, we investigated the effect of the high school background on the university trajectories of both male and female students. In general, the results show that the high school background is an active factor influencing each career step, from the decision to enroll at university (Ballarino and

Panichella, 2016), as shown in Chapter 1, to the transition to the master's level. There has not been a significant narrowing in the gap between students who attended scientific and humanistic "licei" and the others, though allowing students from each track to enroll at university around 50 years ago. Further results show that the high school background differently affects university careers of males and females. The difference in the BA $\rightarrow$ MA transition probabilities tend to favor male students almost always. Additionally, this gap favoring male students is more pronounced among those with a non-liceo background.

Limitations and Future Research

In this section, before considering future developments, it is essential to stress the limitations and assumptions in this work. As for the limitations, the reader must be aware that:

- the lack of information about the socioeconomic background for the oldest student cohorts represents an important drawback. The recent linkage between ANS-U and INVALSI allowed us to gain knowledge of students' socioeconomic status. However, this is only available for one cohort at the moment.
- we have no information about students enrolling or moving to foreign universities. This does not strongly affect the study of student mobility analysis since only a small fraction of students from the South choose to enroll abroad. Nonetheless, this lack of information slightly leads, for instance, to underestimating the overall university enrollment rate at the bachelor's and master's levels.
- time homogeneity in the analysis of student transitions in Chapter 4 is a strong assumption since dropout or graduation times could largely differ based on students' profiles.
- information about part-time could help us explain delays in their university careers.

As for future developments, we distinguish between methodological improvement, data enrichment, and further research questions:

- the recent linkage between ANS-U and INVALSI has opened a large variety of research possibilities. The main objective is to have a deeper understanding of the effects of socioeconomic status on academic outcomes:
  - the first idea is to extend the study in Chapter 1 to include the other types of high schools to get a more complete view of the effect of school background

on university careers. To this aim, the multi-level propensity score should be refined to balance the characteristics of more than two groups. An idea could be to estimate the propensity scores through multinomial models instead of logistic ones.

  - the second idea is to quantify the effect of socioeconomic status on the mobility choices of southern students. As it is reasonable to imagine, we expect that students from lower socioeconomic classes are less likely to move to other regions. As already done in Chapter 1, it could be convenient to apply statistical methods to reduce the imbalance in the socioeconomic status of stayers and movers.

- Given the intricate net of factors influencing educational outcomes, it could be helpful to consider an inter-sectional approach to studying educational differences. This means understanding the complexities potentially held by those possessing multidimensional and intersecting identities such as race, gender, social class, nationality, etc. This approach is introduced in the final chapter of this thesis, where results show how the probabilities of university persistence at different levels vary according to three factors that interact with each other: gender, the field of study, and the type of high school attended.

- The analysis of students' university careers could be investigated by focusing on the variations over time of the transition rates to university controlling for the field of study. In fact, in the last few years, some fields of study have seen a substantial reduction in enrolments. Mediation analysis could be a possible way to investigate this issue.

List of acronyms

Acronyms	**Description**
AMR	Adjusted Migration Rate
ANS	Anagrafe Nazionale Studenti
BA's	Bachelor's degree
CFU	Crediti Formativi Universitari
EHEA	European Higher Education Area
ESCS	Economic, Social and Cultural Status index
HS	High school
ICTs	Information and Communication Technologies
INVALSI	Italian National Evaluation Institute for the School System
ISCED	International Standard Classification of Education
MA's	Master's degree
MPSM	Multi-level Propensity Score Matching
MR	Migration Rate
MUR	Italian Ministry of University and Research
OD	Origin-Destination matrix
PISA	Programme for International Student Assessment
PSM	Propensity Score Matching
STEM	Science, Technology, Engineering, and Mathematics

References

Abramo, G., D'Angelo, C. A., and Rosati, F. (2016). The north-south divide in the italian higher education system. *Scientometrics*, 109(3):2093–2117.

Agasisti, T. and Vittadini, G. (2012). Regional economic disparities as determinants of student's achievement in italy. *Research in Applied Economics*, 4(2):33.

Agresti, A. (2003). *Categorical data analysis*. John Wiley & Sons.

Aina, C., Baici, E., Casalone, G., and Pastore, F. (2018). The economics of university dropouts and delayed graduation: a survey.

Aina, C. and Casalone, G. (2020). Early labor market outcomes of university graduates: Does time to degree matter? *Socio-Economic Planning Sciences*, 71:100822.

Argentin, G. and Triventi, M. (2011). Social inequality in higher education and labour market in a period of institutional reforms: Italy, 1992–2007. *Higher education*, 61(3):309–323.

Arpino, B. and Cannas, M. (2016). Propensity score matching with clustered data. an application to the estimation of the impact of caesarean section on the apgar score. *Statistics in medicine*, 35(12):2074–2091.

Arpino, B. and Mealli, F. (2011). The specification of the propensity score in multilevel observational studies. *Computational Statistics & Data Analysis*, 55(4):1770–1780.

Attanasio, M., Boscaino, G., Capursi, V., and Plaia, A. (2013). Can the students' career be helpful in predicting an increase in universities income? In *Statistical Models for Data Analysis*, pages 9–16. Springer.

Attanasio, M., Enea, M., et al. (2019). La mobilità degli studenti universitari nell'ultimo decennio in italia. *UNIVERSALE PAPERBACKS IL MULINO*, pages 43–58.

Attanasio, M. and Porcu, M. (2021). I grandi numeri dell'istruzione secondaria e terziaria. In *Rapporto sulla popolazione. L'Italia e le sfide della demografia (pp. 163-184). Bologna : UNIVERSALE PAPERBACKS IL MULINO 793.*

Attanasio, M. and Priulla, A. (2020). Chi rimane e chi se ne va? un'analisi statistica della mobilità universitaria dal mezzogiorno d'italia. *In: Massimo Attanasio;Giancarlo Ragozini;Mariano Porcu. "Verso Nord. Le nuove e vecchie rotte delle migrazioni universitarie". ISBN: 9788835105626.*

Attanasio, M., Priulla, A., Albano, A., and Enea, M. (2018). Analisi delle carriere universitarie nelle lauree scientifiche di base in italia nell'ultimo decennio. *INDUZIONI*, pages 37–66.

Austin, P. C. (2011a). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate behavioral research*, 46(3):399–424.

Austin, P. C. (2011b). Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharmaceutical statistics*, 10(2):150–161.

Ballarino, G. and Panichella, N. (2016). Social stratification, secondary school tracking and university enrolment in italy. *Contemporary Social Science*, 11(2-3):169–182.

Ballarino, G. and Panichella, N. (2021). *Sociologia dell'istruzione*. Il mulino.

Ballarino, G., Panichella, N., and Triventi, M. (2014). School expansion and uneven modernization. comparing educational inequality in northern and southern italy. *Research in Social Stratification and Mobility*, 36:69–86.

Barone, C. (2011). Some things never change: Gender segregation in higher education across eight nations and three decades. *Sociology of education*, 84(2):157–176.

Barone, C. and Assirelli, G. (2020). Gender segregation in higher education: an empirical test of seven explanations. *Higher Education*, 79(1):55–78.

Barone, C., Schizzerotto, A., Assirelli, G., and Abbiati, G. (2019). Nudging gender desegregation: A field experiment on the causal effect of information barriers on gender inequalities in higher education. *European Societies*, 21(3):356–377.

Barone, C., Triventi, M., and Assirelli, G. (2018). Explaining social inequalities in access to university: a test of rational choice mechanisms in italy. *European Sociological Review*, 34(5):554–569.

Becker, R. (2003). Educational expansion and persistent inequalities of education: Utilizing subjective expected utility theory to explain increasing participation rates in upper secondary school in the federal republic of germany. *European sociological review*, 19(1):1–24.

Belloc, F., Maruotti, A., and Petrella, L. (2010). University drop-out: an italian experience. *Higher education*, 60(2):127–138.

Bernardi, F. and Ballarino, G. (2014). Participation, equality of opportunity and returns to tertiary education in contemporary europe. *European societies*, 16(3):422–442.

Betts, M. G., Forbes, G. J., and Diamond, A. W. (2007). Thresholds in songbird occurrence in relation to landscape structure. *Conservation Biology*, 21(4):1046–1058.

Blackburn, H. (2017). The status of women in stem in higher education: A review of the literature 2007–2017. *Science & Technology Libraries*, 36(3):235–273.

Borgna, C. and Struffolino, E. (2017). Pushed or pulled? girls and boys facing early school leaving risk in italy. *Social Science Research*, 61:298–313.

Boscaino, G., Sottile, G., and Adelfio, G. (2022). Migration and students' performance: detecting geographical differences following a curves clustering approach. *Journal of Applied Statistics*, 49(4):1018–1032.

Boudon, R. (1974). Education, opportunity, and social inequality: Changing prospects in western society.

Bratti, M., Checchi, D., and de Blasio, G. (2008). Does the expansion of higher education increase the equality of educational opportunities? Evidence from Italy. Technical report.

Bratti, M., Checchi, D., and Filippin, A. (2007). Geographical differences in italian students' mathematical competencies: Evidence from pisa 2003. *Giornale degli Economisti e Annali di Economia*, pages 299–333.

Breen, R., Luijkx, R., Müller, W., and Pollak, R. (2009). Nonpersistent inequality in educational attainment: Evidence from eight european countries. *American journal of sociology*, 114(5):1475–1521.

Briggs, A. R., Clark, J., and Hall, I. (2012). Building bridges: understanding student transition to university. *Quality in higher education*, 18(1):3–21.

Buchmann, C., DiPrete, T. A., and McDaniel, A. (2008). Gender inequalities in education. *Annu. Rev. Sociol*, 34:319–337.

Cappellari, L. (2004). High school types, academic performance and early labor market outcomes. *Academic Performance and Early Labour Market Outcomes (March 2004).*

Caprile, M., Palmén, R., Sanz, P., and Dente, G. (2015). Encouraging stem studies labour market situation and comparison of practices targeted at young people in different member states. *Policy Department A*, 12.

Cascella, C., Giberti, C., and Bolondi, G. (2020). An analysis of differential item functioning on invalsi tests, designed to explore gender gap in mathematical tasks. *Studies in Educational Evaluation*, 64:100819.

Charles, M. and Bradley, K. (2002). Equal but separate? a cross-national study of sex segregation in higher education. *American Sociological Review*, pages 573–599.

Checchi, D. et al. (2003). The italian educational system: family background and social stratification. *Monitoring Italy, ISAE, Rome.*

Checchi, D. and Flabbi, L. (2007). Intergenerational mobility and schooling decisions in germany and italy: The impact of secondary school tracks.

Checchi, D. and Peragine, V. (2005). Regional disparities and inequality of opportunity: the case of italy.

Checchi, D. and Peragine, V. (2010). Inequality of opportunity in italy. *The Journal of Economic Inequality*, 8(4):429–450.

Cheryan, S. (2012). Understanding the paradox in math-related fields: Why do some gender gaps remain while others do not? *Sex roles*, 66(3):184–190.

Cheryan, S., Plaut, V. C., Davies, P. G., and Steele, C. M. (2009). Ambient belonging: how stereotypical cues impact gender participation in computer science. *Journal of personality and social psychology*, 97(6):1045.

Cheryan, S., Ziegler, S. A., Montoya, A. K., and Jiang, L. (2017). Why are some stem fields more gender-balanced than others? *Psychological bulletin*, 143(1):1.

Clancy, P. and Goastellec, G. (2007). Exploring access and equity in higher education: Policy and performance in a comparative perspective. *Higher Education Quarterly*, 61(2):136–154.

Clark Blickenstaff, J. (2005). Women and science careers: leaky pipeline or gender filter? *Gender and education*, 17(4):369–386.

Clerici, R., Giraldo, A., and Meggiolaro, S. (2015). The determinants of academic outcomes in a competing risks approach: evidence from italy. *Studies in Higher Education*, 40(9):1535–1549.

Cochran, W. G. and Rubin, D. B. (1973). Controlling bias in observational studies: A review. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 417–446.

Colombo, M. and Salmieri, L. (2020). Gender and education in italy. pages 72–89.

Columbu, S., Porcu, M., Primerano, I., Sulis, I., and Vitale, M. P. (2021a). Geography of italian student mobility: a network analysis approach. *Socio-Economic Planning Sciences*, 73:100918.

Columbu, S., Porcu, M., and Sulis, I. (2021b). University choice and the attractiveness of the study area: Insights on the differences amongst degree programmes in italy based on generalised mixed-effect models. *Socio-Economic Planning Sciences*, 74:100926.

Contini, D., Cugnata, F., and Scagni, A. (2018). Social selection in higher education. enrolment, dropout and timely degree attainment in italy. *Higher Education*, 75(5):785–808.

Contini, D., Di Tommaso, M. L., and Mendolia, S. (2017). The gender gap in mathematics achievement: Evidence from italian data. *Economics of Education Review*, 58:32–42.

Contini, D. and Scagni, A. (2013). Social-origin inequalities in educational careers in italy. *Determined to succeed*, pages 149–227.

Contini, D. and Triventi, M. (2016). Between formal openness and stratification in secondary education: Implications for social inequalities in italy. In *Models of Secondary Education and Social Inequality*. Edward Elgar Publishing.

Corbett, C. and Hill, C. (2015). *Solving the Equation: The Variables for Women's Success in Engineering and Computing.* ERIC.

Correll, S. J. (2001). Gender and the career choice process: The role of biased self-assessments. *American journal of Sociology*, 106(6):1691–1730.

Crosnoe, R., Riegle-Crumb, C., Field, S., Frank, K., and Muller, C. (2008). Peer group contexts of girls' and boys' academic experiences. *Child development*, 79(1):139–155.

Dal Bianco, A., Ricciari, V., and Spairani, A. (2010). La mobilità degli studenti in italia: un'analisi empirica. *La mobilità degli studenti in Italia*, pages 1000–1021.

D'Angelo, N. and Priulla, A. (2020). Estimating the number of changepoints in segmented regression models. *DSEAS working papers*, IV.

De Vita, L. and Giancola, O. (2017). Between education and employment: Women's trajectories in stem fields. *Polis*, 31(1):45–72.

Dotti, N. F., Fratesi, U., Lenzi, C., and Percoco, M. (2013). Local labour markets and the interregional mobility of italian university students. *Spatial Economic Analysis*, 8(4):443–468.

Duru-Bellat, M., Kieffer, A., and Reimer, D. (2008). Patterns of social inequalities in access to higher education in france and germany. *International journal of comparative sociology*, 49(4-5):347–368.

D'Agostino, A., Ghellini, G., and Longobardi, S. (2019). Out-migration of university enrolment: the mobility behaviour of italian students. *International Journal of Manpower*.

Eccles, J. S. (2007). *Where Are All the Women? Gender Differences in Participation in Physical Science and Engineering.* American Psychological Association.

Eddy, S. (2014). More than half of biology majors are women, yet gender gaps remain in science classrooms. *Science Daily. Arizona State University.*

Enea, M. (2016). From south to north? mobility of southern italian students at the transition from the first to the second level university degree. In *Convegno della Società Italiana di Statistica*, pages 239–249. Springer.

Enea, M. and Attanasio, M. (2020). Gender differences in italian stem degree courses: a discrete-time competing-risks model. In *SIS 2020*, pages 385–390. Pearson.

Fan, S. and Ritz, J. (2014). International views of stem education. *PATT-28 Research into Technological and Engineering Literacy Core Connections*, pages 7–14.

Ferrão, M. and Almeida, L. (2019). Differential effect of university entrance score on first-year students' academic performance in portugal. *Assessment & Evaluation in Higher Education*, 44(4):610–622.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., and Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 111(23):8410–8415.

Gabay-Egozi, L., Shavit, Y., and Yaish, M. (2015). Gender differences in fields of study: the role of significant others and rational choice motivations. *European Sociological Review*, 31(3):284–297.

Ganley, C. M., George, C. E., Cimpian, J. R., and Makowski, M. B. (2018). Gender equity in college majors: Looking beyond the stem/non-stem dichotomy for answers regarding female participation. *American Educational Research Journal*, 55(3):453–487.

García-Peñalvo, F. J., García-Holgado, A., Dominguez, A., and Pascual, J. (2022). Women in stem in higher education: Good practices of attraction, access and retainment in higher education.

Genova, V. G., Tumminello, M., Aiello, F., and Attanasio, M. (2021). A network analysis of student mobility patterns from high school to master's. *Statistical Methods & Applications*, 30(5):1445–1464.

Genova, V. G., Tumminello, M., Enea, M., Aiello, F., and Attanasio, M. (2019). Student mobility in higher education: Sicilian outflow network and chain migrations. *Electronic Journal of Applied Statistical Analysis*, 12(4):774–800.

Gentili, C. (2017). “time out” for classical studies? the future of italian liceo classico in the 4.0 world. 33:127—-143.

Gerber, T. P. and Schaefer, D. R. (2004). Horizontal stratification of higher education in russia: Trends, gender differences, and labor market outcomes. *Sociology of Education*, 77(1):32–59.

Geven, K., Skopek, J., and Triventi, M. (2018). How to increase phd completion rates? an impact evaluation of two reforms in a selective graduate school, 1976–2012. *Research in higher education*, 59(5):529–552.

Giambona, F., Porcu, M., and Sulis, I. (2017). Students mobility: Assessing the determinants of attractiveness across competing territorial areas. *Social indicators research*, 133(3):1105–1132.

Giancola, O., Salmieri, L., et al. (2020). Family background, school-track and macro-area: The complex chains of education inequalities in italy. Technical report.

Gibbons, S. and Vignoles, A. (2012). Geography, choice and participation in higher education in england. *Regional science and urban economics*, 42(1-2):98–113.

Giofrè, D., Cornoldi, C., Martini, A., and Toffalini, E. (2020). A population-level analysis of the gender gap in mathematics: Results on over 13 million children using the invalsi dataset. *Intelligence*, 81:101467.

Giudici, C., Trappolini, E., and Vicari, D. (2021). The academic performance of students with a migrant background: evidence from a cohort enrolled at sapienza university of rome. *Genus*, 77(1):1–23.

Griffith, A. L. (2010). Persistence of women and minorities in stem field majors: Is it the school that matters? *Economics of Education Review*, 29(6):911–922.

Hadjar, A. and Becker, R. (2009). Educational expansion: Expected and unexpected consequences. *Expected and Unexpected Consequences of the Educational Expansion in Europe and the US. Bern: Haupt*, pages 9–23.

Hagan, J. E., Amoaddai, S., Lawer, V. T., and Atteh, E. (2020). Students’ perception towards mathematics and its effects on academic performance. *Asian Journal of Education and Social Studies*, 8(1):8–14.

Hall, R. M. and Sandler, B. R. (1982). The classroom climate: A chilly one for women?. *ERIC*.

Ho, D. E., Imai, K., King, G., and Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political analysis*, 15(3):199–236.

Hocking, R. R. (1976). A biometrics invited paper. the analysis and selection of variables in linear regression. *Biometrics*, 32(1):1–49.

Iammarino, S. and Marinelli, E. (2015). Education–job (mis) match and interregional migration: Italian university graduates' transition to work. *Regional Studies*, 49(5):866–882.

Impicciatore, R. and Tosi, F. (2019). Student mobility in italy: The increasing role of family background during the expansion of higher education supply. *Research in Social Stratification and Mobility*, 62:100409.

Jackson, C. (2007). Multi-state modelling with r: the msm package. *Cambridge, UK*, pages 1–53.

Jackson, M. (2013). *Determined to succeed?: Performance versus choice in educational attainment*. Stanford University Press.

Joensen, J. S. and Nielsen, H. S. (2009). Is there a causal effect of high school math on labor market outcomes? *Journal of Human Resources*, 44(1):171–198.

Keele, L., Lenard, M., and Page, L. (2021). Matching methods for clustered observational studies in education. *Journal of Research on Educational Effectiveness*, 14(3):696–725.

Kim, H.-J., Fay, M. P., Feuer, E. J., and Midthune, D. N. (2000). Permutation tests for joinpoint regression with applications to cancer rates. *Statistics in medicine*, 19(3):335–351.

Kromydas, T. (2017). Rethinking higher education and its relationship with social inequalities: past knowledge, present state and future potential. *Palgrave communications*, 3(1):1–12.

Lauer, S., Momsen, J., Offerdahl, E., Kryjevskaia, M., Christensen, W., and Montplaisir, L. (2013). Stereotyped: Investigating gender in introductory science courses. *CBE—Life Sciences Education*, 12(1):30–38.

Lee, V. E. and Ready, D. D. (2009). Us high school curriculum: Three phases of contemporary research and reform. *The Future of Children*, pages 135–156.

Leemann, R. J. (2010). Gender inequalities in transnational academic mobility and the ideal type of academic entrepreneur. *Discourse: Studies in the Cultural Politics of Education*, 31(5):605–625.

Legewie, J. and DiPrete, T. A. (2012). School context and the gender gap in educational achievement. *American Sociological Review*, 77(3):463–485.

Li, K., Zhang, P., Hu, B. Y., Burchinal, M. R., Fan, X., and Qin, J. (2019). Testing the 'thresholds' of preschool education quality on child outcomes in china. *Early Childhood Research Quarterly*, 47:445–456.

Livi Bacci, M. (2015). Mezzogiorno, italia, europa: Strategie di sviluppo per uscire dalla crisi. *Quaderno SVIMEZ*, 45:55–58.

Lodi, M. (1982). *Guida al mestiere di maestro*. Editori riuniti.

Lucas, S. R. (2001). Effectively maintained inequality: Education transitions, track mobility, and social background effects. *American journal of sociology*, 106(6):1642–1690.

Macarie, F. C. and Moldovan, O. (2015). Horizontal and vertical gender segregation in higher education: Eu 28 under scrutiny. *Managerial Challenges of the Contemporary Society. Proceedings*, 8(1):162.

Maltese, A. V. and Cooper, C. S. (2017). Stem pathways: Do men and women differ in why they enter and exit? *AERA open*, 3(3):2332858417727276.

Mariani, V. and Torrini, R. (2022). Il sistema universitario: Un confronto tra centro-nord e mezzogiorno [the geographic divides of the italian higher education system]. *Bank of Italy Occasional Paper*, (675).

Mega, C., Ronconi, L., and De Beni, R. (2014). What makes a good student? how emotions, self-regulated learning, and motivation contribute to academic achievement. *Journal of educational psychology*, 106(1):121.

Miller, D. I. and Wai, J. (2015). The bachelor's to ph. d. stem pipeline no longer leaks more women than men: A 30-year analysis. *Frontiers in psychology*, 6:37.

Miller, S. and Hughes, D. (2017). The quant crunch: How the demand for data science skills is disrupting the job market. *Burning Glass Technologies*.

Mishra, S. (2020). Social networks, social capital, social support and academic success in higher education: A systematic review with a special focus on 'underrepresented'students. *Educational Research Review*, 29:100307.

MOBYSU.IT (2017). *Database MOBYSU.IT, Mobilità degli studi universitari italiani, Research Protocol MUR - Universities of Cagliari, Palermo, Siena, Torino, Sassari, Firenze, Cattolica and Napoli Federico II, Scientific Coordinator Massimo Attanasio (UNIPA), Data Source ANS-MUR/CINECA*.

Mostafa, T. (2019). Why don't more girls choose to pursue a science career? *OECD*.

Muggeo, V. M. (2003). Estimating regression models with unknown break-points. *Statistics in medicine*, 22(19):3055–3071.

Muggeo, V. M. et al. (2008). Segmented: an r package to fit regression models with broken-line relationships. *R news*, 8(1):20–25.

National Academies of Sciences, E., Medicine, et al. (2018). *Assessing and responding to the growth of computer science undergraduate enrollments*. National Academies Press.

Neugebauer, M., Reimer, D., Schindler, S., and Stocké, V. (2013). Inequality in transitions to secondary school and tertiary education in germany. *Determined to succeed*, pages 56–88.

Nicholas, J., Poladian, L., Mack, J., and Wilson, R. (2015). Mathematics preparation for university: entry, pathways and impact on performance in first year science and mathematics subjects. *International Journal of Innovation in Science and Mathematics Education*, 23(1).

Niss, M. (1994). Mathematics in society. *Didactics of mathematics as a scientific discipline*, 13:367–378.

OECD (2021). *Education at a Glance 2021*.

OECD (2022). *Education at a Glance 2022*.

Panichella, N. (2013). Migration strategies and occupational outcomes of southern italian graduates. *Journal of Modern Italian Studies*, 18(1):72–89.

Panichella, N. and Triventi, M. (2014). Social inequalities in the choice of secondary school: long-term trends during educational expansion and reforms in italy. *European Societies*, 16(5):666–693.

Pensiero, N., Giancola, O., and Barone, C. (2019). Socioeconomic inequality and student outcomes in italy. In *Socioeconomic Inequality and Student Outcomes*, pages 81–94. Springer.

Petolicchio, A. (2016). La riforma del liceo scientifico: il curricolo scienze applicate. *FORMAZIONE & INSEGNAMENTO. Rivista internazionale di Scienze dell'educazione e della formazione*, 14(2):187–194.

Pimentel, S. D., Page, L. C., Lenard, M., and Keele, L. (2018). Optimal multilevel matching using network flows: An application to a summer reading intervention. *The Annals of Applied Statistics*, 12(3):1479–1505.

Poulsen, S. E. (2019). The effect of additional math in high school on college success. *The Mathematics Educator*, 28(2).

Priulla, A., D'Angelo, N., and Attanasio, M. (2021). An analysis of italian university students' performance through segmented regression models: gender differences in stem courses. *Genus*, 77(1):1–20.

Putter, H., Fiocco, M., and Geskus, R. B. (2007). Tutorial in biostatistics: competing risks and multi-state models. *Statistics in medicine*, 26(11):2389–2430.

Ricci, R. (2010). The economic, social, and cultural background: a continuous index for the italian students of the fifth grade.

Rickles, J. H. and Seltzer, M. (2014). A two-stage propensity score matching strategy for treatment effect estimation in a multisite observational study. *Journal of Educational and Behavioral Statistics*, 39(6):612–636.

Romito, M., Gerosa, T., Visentin, M., and Maria, G. (2020). Gender differences in higher education choices. italian girls in the corner? *The Education of Gender The Gender of Education*, page 61.

Rosén, M., Gustafsson, J.-E., and Hansen, K. Y. (2013). Influences of early home factors on later achievement in reading, math and science: An analysis of the swedish data from pirls and timss 2011. In *IEA International Research Conference*, volume 5.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.

Sadler, P. M., Sonnert, G., Hazari, Z., and Tai, R. (2012). Stability and volatility of stem career interest in high school: A gender study. *Science education*, 96(3):411–427.

Salanova, M., Schaufeli, W., Martínez, I., and Bresó, E. (2010). How obstacles and facilitators predict academic performance: The mediating role of study burnout and engagement. *Anxiety, stress & coping*, 23(1):53–70.

Salmieri, L. and Giancola, O. (2020). Gender differences in tertiary educational attainment and the intergenerational transmission of cultural capital in italy. *The Education of Gender The Gender of Education*, page 77.

Santelli, F., Ragozini, G., and Vitale, M. P. (2022). Assessing the effects of local contexts on the mobility choices of university students in campania region in italy. *Genus*, 78(1):1–25.

Santelli, F., Scolorato, C., and Ragozini, G. (2019). On the determinants of student mobility in an interregional perspective: A focus on campania region. *Statistica Applicata-Italian Journal of Applied Statistics*, (1):119–142.

Schizzerotto, A. and Barone, C. (2006). *Sociologia dell'istruzione*. Il mulino.

Schneider, T. (2008). Social inequality in educational participation in the german school system in a longitudinal perspective: Pathways into and out of the most prestigious school track. *European Sociological Review*, 24(4):511–526.

Shavit, Y. and Blossfeld, H.-P. (1993). *Persistent Inequality: Changing Educational Attainment in Thirteen Countries. Social Inequality Series.* ERIC.

Sherman, J. (1980). Mathematics, spatial visualization, and related factors: Changes in girls and boys, grades 8–11. *Journal of Educational psychology*, 72(4):476.

Simon, R. M. (2010). Gender differences in knowledge and attitude towards biotechnology. *Public Understanding of Science*, 19(6):642–653.

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of educational research*, 75(3):417–453.

Stuart, E. A., King, G., Imai, K., and Ho, D. (2011). Matchit: nonparametric preprocessing for parametric causal inference. *Journal of statistical software*.

SVIMEZ (2014). *Rapporto sull'Economia del Mezzogiorno*. Il Mulino.

Tandrayen-Ragoobur, V. and Gokulsing, D. (2021). Gender gap in stem education and career choices: what matters? *Journal of Applied Research in Higher Education*.

Triventi, M. (2010). Something changes, something not. long-term trends in gender segregation of fields of study in italy. *Italian Journal of Sociology of education*, 2(2).

Triventi, M. (2013). The role of higher education stratification in the reproduction of social inequality in the labor market. *Research in Social Stratification and Mobility*, 32:45–63.

Trusty, J. and Niles, S. G. (2003). High-school math courses and completion of the bachelor's degree. *Professional School Counseling*, pages 99–107.

UK-Parliament (2020). Science and technology committee–second report: higher education in science, technology, engineering and mathematics (stem) subjects.

Ulm, K. (1991). A statistical method for assessing a threshold in epidemiological studies. *Statistics in medicine*, 10(3):341–349.

Ulriksen, L., Madsen, L. M., and Holmegaard, H. T. (2015). Why do students in stem higher education programmes drop/opt out?–explanations offered from research. In *Understanding student participation and choice in science and technology education*, pages 203–217. Springer.

UNESCO (2013). International standard classification of education: Fields of education and training 2013.

Vergolini, L. and Vlach, E. (2017). Family background and educational path of italian graduates. *Higher Education*, 73(2):245–259.

Vettori, G., Vezzani, C., Bigozzi, L., and Pinto, G. (2021). Patterns of academic performances and conceptions of learning in on-time, delayed, and dropout university participants: a cohort retrospective study. *Educational Psychology*, pages 1–17.

Vittorietti, M., Giambalvo, O., and Aiello, F. (2019). A mobility analysis of the occupational status of the graduates of the university of palermo in an economic crisis context. *Electronic Journal of Applied Statistical Analysis*, 12(4):846–869.

Vittorietti, M., Giambalvo, O., Genova, V. G., and Aiello, F. (2022). A new measure for the attitude to mobility of italian students and graduates: a topological data analysis approach. *Statistical Methods & Applications*, pages 1–35.

Wang, M.-T. and Degol, J. (2013). Motivational pathways to stem career choices: Using expectancy–value perspective to understand individual and gender differences in stem fields. *Developmental Review*, 33(4):304–340.

Wang, X. (2013). Why students choose stem majors: Motivation, high school learning, and postsecondary context of support. *American Educational Research Journal*, 50(5):1081–1121.

Watkins, J. and Mazur, E. (2013). Retaining students in science, technology, engineering, and mathematics (stem) majors. *Journal of College Science Teaching*, 42(5):36–41.

Weeden, K. A., Gelbgiser, D., and Morgan, S. L. (2020). Pipeline dreams: Occupational plans and gender differences in stem major persistence and completion. *Sociology of Education*, 93(4):297–314.

Wintre, M. G. and Bowers, C. D. (2007). Predictors of persistence to graduation: Extending a model and data on the transition to university model. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, 39(3):220.

CRediT Author Statement

- **Chapter 1** - ***Andrea Priulla:*** conceptualization, software, formal analysis, investigation, data curation, writing - original draft, visualization. ***Martina Vittorietti:*** methodology, software, investigation, writing - review and editing, supervision. ***Massimo Attanasio:*** conceptualization, supervision, writing - review, project administration, and funding acquisition.

- **Chapter 2** - ***Andrea Priulla:*** conceptualization, software, formal analysis, investigation, data curation, writing - original draft, visualization. ***Massimo Attanasio:*** conceptualization, supervision, project administration, funding acquisition.

- **Chapter 3** - ***Andrea Priulla:*** conceptualization, methodology, software, investigation, formal analysis, data curation, writing - original draft, visualization. ***Nicoletta D'Angeo:*** conceptualization, methodology, software, writing - original draft. ***Massimo Attanasio:*** conceptualization, supervision, project administration, funding acquisition.

- **Chapter 4** - ***Andrea Priulla:*** conceptualization, methodology, software, investigation, formal analysis, data curation, writing - original draft, visualization. ***Massimo Attanasio:*** supervision, writing - review, project administration, and funding acquisition.

Funding

This thesis has been supported from Italian Ministerial grant PRIN 2017 "From high school to job placement: micro-data life course analysis of university student mobility and its impact on the Italian North-South divide.", n. 2017HBTK5P - CUP B78D19000180001.