Università degli Studi di Palermo

AREA QUALITÀ, PROGRAMMAZIONE E SUPPORTO STRATEGICO
SETTORE STRATEGIA PER LA RICERCA
U. O. DOTTORATI

Corso di Dottorato in Scienze Economiche e Statistiche
Dipartimento di Scienze Economiche, Aziendali e
Statistiche Settore Scientifico Disciplinare: SECS-S/05

SOME STATISTICAL METHODS TO ANALYSE STUDENTS' TRANSITION TO UNIVERSITY AND FACULTY'S CAREER ADVANCEMENTS

IL DOTTORE
VINCENZO FALCO

IL COORDINATORE
VITO M. R. MUGGEO

IL TUTOR
MASSIMO ATTANASIO

IL CO-TUTOR
MARTINA VITTORIETTI

CICLO XXXVI
ANNO CONSEGUIMENTO TITOLO: 2024

≪*Ero così ottimista da sperare che la morte*
avrebbe posto fine alle mie sofferenze.
Sono morta di ottimismo.≫
Viola Di Grado

A Dorotea, la mia anima

Contents

Acknowledgements

Come i monsoni mi sono sentito in tempesta e come i deserti mi sono sentito senza acqua, ghiacciato, eppure mai solo. Voglio solo dire grazie a chi mi ha accompagnato e sostenuto lungo questo percorso, voglio dire grazie a chi ha lasciato il segno nella mia vita, voglio dire grazie a chi e tutto ciò che fa parte del passato oppure di un'altra galassia e non è più nella mia.

Desidero esprimere la mia gratitudine, innanzitutto, al mio tutor Massimo Attanasio e alla mia co-tutor Martina Vittorietti per la loro inestimabile pazienza e il loro aiuto costante. Un ringraziamento speciale va alla professoressa Ornella Giambalvo che mi ha sempre fornito generosamente conoscenze e affetto. Infine, voglio ringraziare i proff. Giovanni Boscaino e Mariangela Sciandra per essere stati presenti anche quando il mio percorso è diventato difficile e tutti hanno preferito voltarmi le spalle e Daniele con cui ho condiviso tutta la mia avventura universitaria.

Voglio ringraziare il dottorato per avermi regalato Cristina, una delle persone migliori che conosca, che mi ha reso migliore, che è diventata parte di me, che mi vizia, mi protegge e mi ama incondizionatamente, che mi ha regalato persone speciali, che realizzerà un sogno che la mia vita ha ucciso perché i sogni e la felicità di Cristina sono anche miei.

Un ringraziamento speciale va alla mia pazzologa Raffaella che mi ha cambiato la vita, che mi conosce meglio di chiunque altro, che mi ha fatto scoprire me stesso al buio e luccicante.

Tutto questo non sarebbe stato possibile senza la mia personale grande bellezza. Mamma e papà che ci sono sempre, per sempre, qualunque cosa accada. Tutto questo non sarebbe stato possibile senza mia sorella Federica che mi ha insegnato ad essere una ragazza di talento, l'amore e l'amicizia, alcuni o forse tutti i significati. Tutto questo non sarebbe stato possibile senza my friend Sara che mi ascolta e che mi dà le pacche sulle spalle e rimane sempre accanto a me. Tutto questo non sarebbe stato possibile senza l'agrumato ricordo di ogni giorno felice sull'erba, delle corse sul mio giardino incantato di sogni, amore e fischietti colorati.

Durante questi anni non è passato un solo giorno, un solo minuto, un solo attimo senza pensare che vorrei morire di ottimismo come te Dorotea. Circondato di sangue e lacrime nella vasca. Il mondo è morto e continua a morire con te, Dorotea. Ho

la tua assenza con me, la amo come amavo te, come amavo ciò che mi rendeva me stesso e non c'è più. Ho il tuo cuore cavo che è diventato il mio anche se è sottoterra perché mi sono sentito sottoterra come te. Se non sono ancora morto di ottimismo come te è solo perché ho ancora voglia di guardare e annusare il profumo delle viole che ci piacevano tanto.

Introduction

Modern university education encompasses different teaching strategies, theoretical frameworks and socio-political processes. It is a complex system influenced by institutional frameworks, historical legacies and external factors. In today's academic language, education is presented as a dynamic, multidimensional field constantly changing and being studied (Barnett, 2015). The pursuit of knowledge in education spans a wide range of fields, approaches and focal points, all of which contribute to our understanding of the complexity of education systems (Labaree, 2010).

Universities have their roots in the centres of learning that emerged in ancient civilisations to foster intellectual inquiry and preserve knowledge (Van Doren, 1991). However, the modern idea of a university first emerged in medieval Europe, when institutions such as the University of Bologna and the University of Paris laid the foundations for academic freedom and intellectual endeavour. Universities were further developed during the Renaissance and Enlightenment, when humanism, rationalism and scientific inquiry were developed.

University education is changing dramatically in the 21st century, driven by globalisation, technological breakthroughs, demographic shifts and changing societal demands. Beyond individual achievement, higher education is essential for wider socio-economic development and cultural enrichment. Policymakers, educators, researchers, and other stakeholders involved in setting the direction of higher education must, therefore, be fully aware of the complexity of university education.

This thesis aims to investigate specific aspects of educational system within this dynamic framework. Throughout the entire educational journey, including university education, individuals go through various steps, more specifically they experience "transitions". Exploring what is the impact of individual differences in responses to educational interventions, socio-economic variables, and environmental conditions during transitions is one of the main objective of this thesis.

In particular, we focus on two different and interrelated transitions: the academic career paths of Italian faculty members and the students' educational paths. Even though these two topics might not seem similar at first, they have some things in common that are worth investigating. Though from different points of view, both topics centre on the essential idea of educational growth and transition: university faculty members' career promotion and students' journeys from high school to university education.

A multitude of factors impact the experiences and paths of both students and faculty members. These variables affect pupils in addition to academic difficulties and consider social, personal, and financial problems (Pascarella and Terenzini, 2005). Conversely, faculty members must negotiate changes in pedagogy, institutional expectations, and professional growth (Bauer and Bennett, 2003). In addition, there are many different institutions, parties, and regulations involved in education systems, giving them a complex nature. As it is a challenge to fully capture alla relevant data, data integration plays a crucial role. The integration of different data sources in the Italian context has made it possible to have microdata at individual level and in aggregated form. This data integration approach allow us to apply appropriate statistical methods and obtain valid results and conclusions.

Statistical analysis plays a crucial role in understanding educational transitions and phenomena. In particular, longitudinal analysis provides important insights into the factors influencing educational outcomes and trajectories by monitoring patterns of change and continuity over time (Bryk and Raudenbush, 2002). The advantage of longitudinal data analysis is capturing the interplay between individual, contextual, and temporal elements that influence educational outcomes and unveiling the mechanisms underlying educational success or failure. Longitudinal data analysis is characterised by its ability to track changes and developments over time. This temporal dimension is crucial for understanding not only what changes are occurring, but also when they are occurring, providing insights into the relationships and time-dependent effects of educational phenomena.

This thesis is divided into three chapters.

Chapter 1 reviews the statistical methods, especially longitudinal data analysis approaches, used to analyse transitions in the context of university education. Transitions, understood as significant jump from one educational stage to another, represent critical moments in individuals' educational journeys and can significantly influence their academic and professional outcomes. Therefore, understanding and analysing such transitions is essential to provide an in-depth view of educational processes. Finally, the challenges and limitations associated with analysing transitions in university education are discussed.

Chapter 2 focuses on academic transitions in Italian universities over the last 20 years, especially in terms of gender differences (from assistant to associate professor and from associate professor to full professor). The data are taken from the Ministry of Universities and Research (MUR) archives. The statistical methodology applied refer to event history analysis. We examine the temporal trends of gender differences in academic transitions in Italian universities over the last 20 years, highlighting how they have evolved and analysing some factors contributing to gender differences in academic transitions. The role of the sociological metaphors of "leaky pipeline" and the "glass ceiling" in the context of Italian universities is also explored, highlighting how these concepts can interpret the gender gap in academic careers.

Chapter 3 examines the transition from high school to university for students in Italy, addressing some of the factors influencing this critical step. We analyse the persistent inequalities in the Italian educational landscape. We build an aggregate database containing data related to school-university transiton for the 2019/2020 academic year. To account for the non-random selection of students in high schools and the potential for selection bias, we compare two balancing methods. Based on the data obtained after balancing, we estimate the effect of some variables affecting the school-university transition.

Chapter 1

Transitions in educational studies

This chapter is based on the work from Falco, V., Attanasio, M. (2024). A review of statistical methods for analysing transitions in higher education. *Work in progress*

Abstract

This chapter explores the concept of transition in university education and reviews the statistical methods used to analyse transitions, focusing on the educational pathways of students and the academic careers of faculty members. Specifically, it then looks at longitudinal data analysis, highlighting its usefulness in capturing the dynamics of educational pathways over time. Finally, the chapter looks at challenges and emerging approaches to data quality, availability and complexity in modelling educational transitions.

1.1 Introduction

The concept of "transition" is through various fields and contexts and has no single meaning. Epidemiological transitions denote structural shifts in disease patterns and health behaviors within populations (Barrett et al., 1998). Similarly, in demography, transitions refer to population structure and dynamic changes. Transitions are relevant in physics, biology, economics, and other fields. In these contexts, transitions are gradual changes over time. In other contexts, e.g. educational studies, a transition denotes a transition from one state to another, indicating a point of "discreet" and change that has an impact on individuals and society (Terenzini, 1999).

The emphasis of this chapter is on educational transitions, focusing specifically on university education. These transitions are complex processes that shape the educational environment. Understanding these transitions is crucial for various reasons. Longitudinal data analysis offers a panoramic view of evolving patterns over time and tracks the same information on the same subjects at multiple points in time. Specifically, longitudinal data analysis consists of the statistical tools and methods used to analyse data collected on the same group of individuals on multiple occasions over time. By examining longitudinal data, researchers can gain insights into the factors influencing progress (Pascarella and Terenzini, 2005). In university education, longitudinal data analysis captures student's progression through various educational steps, from high school to university, and through different career steps (Perry and Smart, 1997). Similarly, on the faculty front, longitudinal data tracks the career trajectories of faculty members from entry-level positions to top roles (Baldridge et al., 1978).

This chapter aims to review the most common statistical methods used to analyse longitudinal data in the context of university education.

The chapter is divided into four sections: after a brief introduction, a review about transitions and statistical modelling of them in university education will be provided. Some final considerations about challenges follow.

1.2 Transitions in university education for students and faculty members

Transitions in university education encompass the movement of individuals, both students and faculty members, into, through, and out of higher education institutions. These transitions are critical junctures that shape educational trajectories, career pathways, and institutional dynamics (Tinto, 1975). Various kinds of transitions can affect students and faculty members in university education.

Moving from high school to university represents a significant transition for students (Tinto, 1975). This transition involves adapting to a new academic environment, navigating the challenges of increased autonomy and academic rigour, and establishing new social networks. Choosing a major or academic pathway is another critical transition for students (Bean and Metzner, 1985). This decision often influences their course of study, career aspirations, and long-term academic and professional goals. Other significant transitions for students during university education concern performance at university in terms of first-year dropout, course transfer, or continuation of studies (De Laet and Rubin, 2008). Pursuing graduate studies or postgraduate education represents a significant transition for students aspiring to advance their academic and professional careers (Scott-Clayton, 2012). This transition involves deepening expertise in a specific field, conducting research, and preparing for careers in academia, industry, or other sectors. It's important to acknowledge that not all high school graduates follow the traditional path to university. For instance, Eurostat data from 2020 shows that throughout Europe, roughly 39% of people between the ages of 20 and 24 were enrolled in postsecondary education, suggesting that a significant proportion of young adults in the region choose not to continue their education after high school (Eurostat, 2020). After high school, those who decide not to go to college may choose to transition by either getting a job right away or looking into other options like apprenticeships or vocational training. Numerous considerations, such as budgetary limitations, individual preferences, professional goals, and societal trends, may impact this choice (Rivera, 2019).

On the other hand, attaining tenure and promotion represents a critical transition for faculty members. These transitions involve demonstrating excellence in teaching, research, and service to the institution, culminating in the award of tenure and advancement in academic rank. Transitioning into leadership and administrative roles presents new challenges and responsibilities for faculty members. This transition involves overseeing academic programs, managing resources, and shaping institutional policies and initiatives. Transitioning into retirement and post-retirement activities marks the culmination of a faculty member's academic career (Scott and Meyer, 1994). This transition involves planning for retirement, pursuing emeritus status, and engaging in continued scholarly pursuits, mentoring, or community service. Understanding the diverse transitions students and faculty members experience in university education is crucial for fostering academic and professional growth, promoting institutional success, and ensuring the continued vitality of higher education (Tinto, 1975).

1.3 Statistical modelling for transitions in university education

In order to capture the complex dynamics of change over time, statistical models are necessary when examining transitions within the context of university education. The most common statistical methods often were born in other contexts, but they are also helpful in university education. While many statistical methods have originated in other fields, their utility extends to university education. This section outlines some of the most common statistical methods employed in this context.

We start with Event History Analysis (EHA), a statistical framework that provides insight into event occurrence and timing throughout time, especially in longitudinal research, before delving into other statistical methodologies for university education.

Event history analysis is a statistical framework that provides insight into the occurrence and timing of events over time, particularly in longitudinal studies (Allison, 1982). EHA is initially rooted in survival analysis, which forms the basis of EHA (Cox, 1972). Survival analysis focuses on understanding the time to occurrence of one or more events of interest (Klein and Moeschberger, 2003). It allows the analysis of longitudinal data where the outcome variable is the time to the occurrence of an event (Cleves et al., 2004) and it also manages censored observations and covariate variables over time (Hougaard, 2000; Allison, 1984). If the event of interest is one, then the single transition model analyses a single change of state, i.e. the transition from an initial to a final state. If the event of interest is multiple, then we consider multiple transitions between them and a potentially complex path. Solutions such as incorporating variables that account for previous states can be used. Multi-state models provide a more rigorous approach to analysing transitions between different states or statuses over time, such as moving from one level of education to another or between different career stages (Aalen et al., 2008). When events of the same type are repeated over time for the same individual, we speak of Multi-episode models. Finally, if the different types of events can prevent or compete with each other to occur first, then competing risk models are considered (Fine and Gray, 1999; Andersen and Keiding, 2012). Graphically:

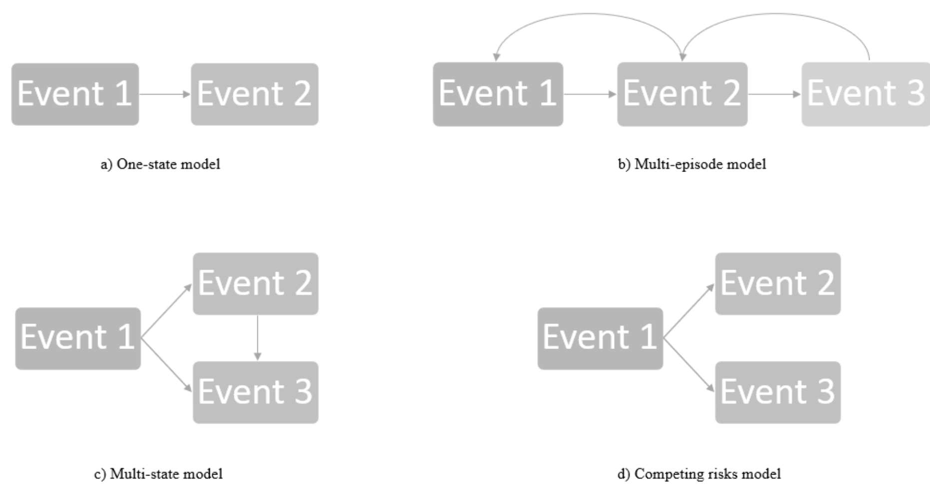

Figure 1.1: Comparison of "EHA" models

These methods are interconnected within the broader framework of EHA, which allows for the modelling and analysis of the timing and occurrence of events within longitudinal data (Hougaard, 2000) and provides deeper insights into the factors influencing educational and career transitions, identifying critical points in individuals' trajectories and exploring the interplay between different events and outcomes over time.

When introducing EHA, it is also important to distinguish between discrete-time and continuous-time models. Discrete-time models treat time as a series of distinct intervals. This approach is particularly useful when data are collected at regular, discrete points in time (e.g. semesters, years, or other predefined periods) and simplifies the treatment of events by aggregating events within these intervals. Continuous-time models, on the other hand, treat time as a continuous variable. This approach allows more precise modelling of the exact timing of events, capturing the nuances of when events occur.

Its versatility finds application in various fields, including sociology and demography, medical and health research, economics, etc. EHA is also a valuable tool for understanding various transitions experienced by students and faculty members within the university setting.

From the students' point of view, the first transition in university education is from high school to university. In Italy, usually, the focus had been on the determinants of students' choices to attend bachelor degree studies outside the region of residence such as field of study (Columbu et al., 2021), high school and microregion of residence (Usala et al., 2023). This suggests that despite advancements in university access, disparities persist based on students' educational pathways. Within university education, another significant transition process involves the geographical mobility of students, both during the school-university transition and potentially when enrolling in a master's degree course. In Italy, this mobility increased and has only one direction: from the South to the Center-North during the school-university transition and the transition to a master's degree course (Attanasio and Priulla, 2020). The relationship between the attitude to the mobility of students enrolled at university and the main socio-demographic variables can be achieved through a combination of descriptive statistical analysis, analysis of group differences, multivariate analysis, and qualitative approaches to understand the factors influencing this attitude. The relationship between the mobility attitude of Italian students enrolled at university and the main socio-demographic variables was also investigated through a new measure, combining tools from graph theory and topological data analysis (Vittorietti et al., 2022). Another point of view regarding territorial mobility concerns students' flows across competing territorial areas which supply university education programs. A study considers a wide range of determinants related to the socio-economic characteristics of the areas as well as resources of the universities in the territories in terms of the variety and quantity of the degree programs there available, financial endowments provided by the Central Government, and services available to students (Giambona et al., 2017). In the first year of university, a common transition is the dropout. Survival analysis techniques were employed to predict student dropout in a large university cohort (Johnson et al., 2018). Prior academic performance, financial aid status along socio-economic lines (Contini and Zotti, 2022), and student engagement significantly influence dropout rates. Other results based on a machine-learning based method highlight the influence of other factors, such as the importance of dedication (part or full-time), and the vulnerability of the students with respect to their age (Rodriguez-Muñiz et al., 2019).

On the faculty front, for instance, models were used to analyse faculty promotion patterns in a research-intensive university (Chen et al., 2019). The study examined factors associated with successful promotion to associate professor and full professor ranks, including publication productivity, grant funding, and teaching evaluations. The findings revealed significant differences in promotion trajectories across disciplines and highlighted the importance of mentorship and research collaboration in facilitating academic transition.

These methodologies offer valuable insights into various transitions experienced by students and faculty members (Smith et al., 2018; Jones and Brown, 2020). By employing these methodologies within the context of university education, researchers can gain a comprehensive understanding of the dynamics of academic trajectories and inform evidence-based policies and interventions aimed at promoting student success and faculty academic transition within higher education institutions (Brown and Williams, 2019; Martinez et al., 2020).

Another class of models is the Growth Curve Models (GCMs). GCMs are versatile and powerful tools for capturing the dynamic patterns of change over time within individual trajectories. These models offer a comprehensive framework to analyse and interpret longitudinal data across various disciplines (Singer and Willett, 2003). These models consider fixed and random effects, allowing for exploring population-average trends and individual variability in trajectories. Like the methods belonging to the EHA, GCMs are also used in various fields, especially in the econometric context (Lancaster, 1990). In the university education context, GCMs are valuable for studying changes in student outcomes over time and analysing faculty research productivity and career trajectories (Raudenbush and Bryk, 2002). These models can be used to examine students' learning trajectories over time, including individual variations in academic outcomes and the effects of educational policies or teaching interventions (Jones et al., 2018; Johnson and Smith, 2020). Additionally, growth curve models can be employed to understand career transitions for faculty members, such as shifts in research focus, teaching methodologies, or administrative responsibilities over their tenure within the university (Brown and Lee, 2017; Smith and Johnson, 2019).

Hidden Markov Models (HMMs) also represent a powerful statistical framework for studying transitions between unobservable states over time, latent states and transitions in dynamic systems. HMMs offer a flexible and nuanced approach to capturing latent processes and understanding complex sequences of events (Jackson et al., 2011). In university education, HMMs have found applications in studying transitions of students and faculty members within academic environments. Research in this area has used HMMs to understand student trajectories, predict dropout rates, and identify factors influencing academic success (Thangarajah et al., 2016). By modelling transitions between states such as active engagement, disengagement, and academic achievement, HMMs provide a nuanced understanding of educational dynamics and contribute to developing effective intervention strategies for improving student outcomes. Furthermore, HMMs can elucidate career progression patterns, workload distribution, and professional development trajectories in faculty transitions. Through analysing transitions between teaching, research, and administrative responsibilities, these models support academic growth and faculty allocation optimisation (Lafferty et al., 2001).

Latent Transition Analysis (LTA) is a statistical method for modelling transitions between unobservable states in categorical variables over time. This approach enables researchers to uncover hidden structures within longitudinal data and understand how individuals move between different latent states (Collins and Lanza, 2010). Latent Transition Analysis is widely used in educational research to study transitions between educational states. In this context, LTA is flexible to model changes in peer victimisation (Lubke and Muthén, 2005). Additionally, LTA has been applied to examine educational pathways and transitions in academic achievement (Wang and Brown, 2009), providing valuable insights of educational' dynamics trajectories. For instance, LCA is used to identify unobserved groups or classes of students with similar behaviour patterns or characteristics. For students, LCA can inform interventions by identifying sub-groups at risk of dropout or academic underperformance. For faculty, LCA can classify researchers into distinct productivity profiles based on publication records (Collins and Lanza, 2010). Recent advances in Latent Transition Analysis involve extensions and refinements to the basic model. One such example concerns an extension of LTA to handle missing data and measurement non-invariance, addressing practical challenges in longitudinal studies (Wang et al., 2018). So, LTA continues to be a valuable tool for unraveling complex patterns of transitions in categorical variables across different domains.

Finally, we provide a summary table for the various methodologies mentioned above, their advantages and disadvantages.

Table 1.1: Summary of statistical methodologies used in university education transitions

Methodology	**Advantages**	**Disadvantages**
Event History Analysis Survival analysis	Captures the timing and occurrence of events Handles censored data Can model multiple transitions Flexible extended models	Requires large sample sizes Complex implementation and interpretion Sensitive to model specification and assumptions (PH)
Growth Curve Models	Captures individual trajectories over time Models both fixed and random effects Versatile applications	Requires advanced statistical knowledge Sensitive to missing data
Hidden Markov Models	Models transitions between unobservable states Handles complex sequence data Flexible and powerful	Computationally intensive Requires large datasets Complex parameter estimation
Latent Transition Analysis	Uncovers hidden structures in categorical data Handles longitudinal transitions Flexible modelling	Complex to set up and interpret Sensitive to model assumptions and missing data

1.4 Challenges

The complexity of transition processes in the context of university education poses several methodological and empirical challenges. These include the quality and availability of data, and the complexity of analysing transitions with appropriate statistical methods.

Firstly, the issue of data availability hinders comprehensive analysis. University education systems vary across countries, making it challenging to collect standardised data. Furthermore, longitudinal data that track individuals' educational trajectories are often scarce or fragmented (Blossfeld and Rohwer, 2002). This limitation restricts the ability to capture the entirety of individuals' educational journeys and may introduce biases in analyses. Data integration from various sources offer promising opportunities to overcome these challenges and broaden our understanding of transitions in university education. Secondly, in contrast to experimental research, social science research involves the problem of data quality and the necessity to "clean" and "adjust" them to use statistical methods.

The dynamic nature of transitions necessitates sophisticated modelling techniques. Traditional statistical methods may not adequately capture the intricacies of educational pathways characterised by multiple transitions, non-linearity, and interdependencies among various factors (Di Prete and Buchmann, 2013). Therefore, there is a growing need for advanced statistical approaches to accommodate these complexities. Recent advancements in statistical methods of longitudinal data analysis offer promising avenues for overcoming some of these challenges (StataCorp, 2019) to understand educational transitions and promote educational equity and attainment.

Also the interpretation of statistical results in the context of educational transitions requires careful consideration of contextual factors and potential confounders. Socio-economic background, educational policies, and institutional characteristics can significantly influence individuals' transition patterns (Van de Werfhorst and Mijs, 2010). Ignoring these contextual factors may lead to biased conclusions and misinterpretations of the underlying dynamics.

In summary, while studying transitions in higher education comes with several difficulties, new statistical techniques bring intriguing chances to get beyond these restrictions and expand our knowledge of educational paths.

In the subsequent chapters of the thesis, we will further explore two specific transitions in university education, focusing on how these processes can be studied and understood by applying advanced statistical methods. We will mainly delve into the transitions of students from high school to university and the transitions of faculty members' academic careers within the university.

Chapter 2

Academic transitions in Italian universities over the last 20 years

This chapter is based on the work from Falco, V., Cuntrera, D., Attanasio, M. (2022). Gender differences in career advancements in Italian universities over the last 20 years. *Genus*, 79 (14)

Abstract

This chapter deals with academic transitions in Italian universities in the last 20 years, particularly in terms of gender differences. Data is taken from the MUR (Ministry of University and Research) archive. In Italy, academic transitions are still much easier for men, even if the gender gap has slowly narrowed in the last decades. The novelty is the analysis through event-history analysis models on the time elapsed to receive a promotion (from assistant to associate professor and from associate to full professor). The event-history analysis applied to academic transitions has revealed that women take, on average, about one and a half more years than men to advance, with some differences among fields of study and macroregions. Furthermore, this gender gap is higher in the first years of the career. Two sociological metaphors used in the gender literature, the "leaky pipeline" and the "glass ceiling", seem to intervene powerfully in the gender gap of Italian universities careers.

2.1 Introduction

The recruitment and selective academic transition of faculty members are essential for the medium and long-term future of academic systems. Gender inequalities in academic careers can be usefully analysed across four dimensions: participation; position; productivity; and recognition (Long and Fox, 1995).

Substantial differences characterize academic systems worldwide, especially in terms of the recruitment process of faculty members and the gender gap remains a crucial issue, as men are more likely and faster in their career than woman. Sexism in academia is experienced at both the systemic and individual levels, often culturally normalized and personally violent yet cloaked beneath veneers of professional activity and working relationships. The silence surrounding sexism in academia is deafening (Teixeira et al., 2018; Thun, 2020). In recent years, numerous studies have explored the gender gap in academic careers, particularly focusing on the reasons behind the slower career progression of women compared to men such as implicit biases and stereotype threat in hindering women's advancement in academia. These psychological phenomena can contribute to the underrepresentation of women in higher academic positions. Significant differences between the sexes can be seen in terms of inequalities in access to professional careers (Kirchmeyer, 2002), in reaching the highest positions and in terms of career speed in reaching those positions ((Forum, 2021), (Commission, 2021)). In general, there are also differences in terms of overall remuneration (Boden Jr., 1999), in salary in universities (Gibelman, 2003; Ginther and Hayes, 1999) and in the corporate and financial world ((Bertrand and Hallock, 2001), (Bertrand et al., 2010), (Elkinawy and Stater, 2011)). Significant disparities have also been highlighted by other studies with respect to credit access (Marlow and Patton, 2005) and public funding (Vossenberg, 2013). In short, even in 2021, in many countries with advanced economies, the gap remains significant despite equalizing gender policies.

Contemporary academic and policy research on academic employment has shown how academics continue to experience gender differences (David and Woodward, 1998). In Europe, only three countries - Iceland, Latvia, and Lithuania - have a higher proportion of female than male researchers in higher education (OECD, 2020). Then, there are fewer women as the standard academic career progresses. In the EU-28 in 2018, women made up 47% of assistant professors, 40% of associate professors and 26% of full professors. At the national level, the proportion of women among full professors ranged from 18% to 51%, exceeding 50% in only one country (Commission et al., 2021). However, since 2013 the proportion of full women professors has increased in almost all European countries, and the increase continues to persist to 2018, albeit slowly. The only exceptions were Hungary and Spain. The largest increases were observed in Latvia, Slovenia, and Romania.

The Italian experience is in line with the European one. In fact, women are under-represented in the academic staff (one-third of the population) and the proportion varies across fields and academic positions (Abramo et al., 2021). The road towards gender equality is extremely slow and non-linear (Gaiaschi and Musumeci, 2020). Their lower scientific productivity cannot explain the lower likelihood that women will be promoted to associate and full professorships in Italian universities. Nor can it explain the negative self-selection of women who are less inclined to apply for promotion (Filandri and Pasqua, 2019).

The goal of this chapter is to study gender differences in academic transition intervals in Italy: to look at how they vary over time; to measure the length of time necessary to get a promotion in the transition from assistant professor to associate professor and from associate professor to full professor; to ask how those transitions are influenced by covariates like macroregion and field of study.

2.2 Background

In general, studies have shown that women progress slower through the academic ranks, that they tend not to attain essential leadership roles, and that they earn less than men in comparable positions (Peterson, 2016; Van den Brink and Benschop, 2012a,b).

The question that naturally arises is which factors, or which combination of factors, could cause this under-representation. Intersectionality is treated in several papers (Collins and Bilge, 2020). Beyond the intersectionality effects, it is common to refer to two sociological methaphors, well known in the literature to describe this gender gap: the "glass ceiling effect" and the "leaky pipeline" metaphors. These metaphors have taken on particular significance in the analysis of academic careers.

- The "glass ceiling effect" was used for the first time by an American writer, Loden, in a 1978 speech, and, in the same year, by two executives, Schriber and Lawrence at Hewlett-Packard. It highlights an invisible barrier that prevent women from climbing the career ladder beyond a certain level (Wirth, 2001; Cotter et al., 2001). In 1986, two journalists, Hymowitz and Schellhardt, commented on the glass ceiling as something that could not be found in any corporate manual or discussed at a business meeting. It was an invisible and unspoken phenomenon that existed for executive-level leadership positions. These positions are reserved for men. Cotter et al. (Cotter et al., 2001) identify four criteria, in scientific literature, that might be used to define a glass ceiling effect: a) in levels of authority; b) in positions in the corporate hierarchy; c) in earnings; and d) in occupation. These criteria are generally framed in terms of "outcomes" rather than for one specific measure. So, they identify the four criteria as: i) a gender (or racial) difference that is not explained by the other job-relevant characteristics of the employee; ii) a gender (or racial) difference that is greater at higher levels than at lower levels of an outcome; iii) gender (or racial) inequality in the chances of advancement into higher levels, not merely the proportions of each gender or race currently at those higher levels; iv) gender (or racial) inequality that increases as a subject moves upwards through a career hierarchy. Wright and Baxter (Baxter and Wright, 2000) give a broad definition of the "glass ceiling", in which the difficulties of the ladder are present and increasing at each step. The latter definition looks closer than the first one to female academic transition in Italian universities.

- The "leaky pipeline" is a metaphor born in the studies of women in science, referring both to the loss of female students in high school and university education and - later on - across the different steps of the career ladder in scientific professions, including (but not exclusively to) academia (Blickenstaff, 2005; Alper, 1993). A system, in this metaphor, is designed to channel something from one place to another. But the system malfunctions because it leaks losing some of what it carries before it reaches the destination. In academia, it would seem that the academic system clogs up rather than leaks: i.e., it lets some through more quickly than others. And often, it fails outright.

Several papers introduce concepts closer to the leaky pipeline ((Alper, 1993; Blickenstaff, 2005)), as "sticky floors" and "mid-level bottlenecks", ((Yap and Konrad, 2009)). These concepts are related to the women's discouragement in approaching the university career because it will be too demanding, precarious, and difficult to be compatible with family work-load ((Bataille et al., 2017)). Interestingly, the concept of "gender blindness" ((Thun, 2020)) referred to the academic organization which does not take into account of the motherhood and female needs. In the Italian case, we do not have studies on the discouragement effect (which affects mostly the PhD's) but it seems that the system is "leaky" along with the entire ladder of the university career.

Other factors range from subtle gender stereotyping, patterns of socialization and upbringing, via issues of family and domestic responsibilities, to the structuring of academic work and career paths (Roth and Sonnert, 2010). There are also demographic cohort factors (current faculty members have been recruited from cohorts where the proportion of women in the universities was much lower than today) (Stewart et al., 2009):

- gender differences in the recruitment, the advancement process, and the workplace (Regner et al., 2019);
- self-selection (personal choices) (Barrett and Barrett, 2011);
- etc.

Other studies indicate that women remain in stagnant career positions or leave their occupations to become full-time caregivers (Stone, 2007) or pursue other careers for many different reasons. Some scholars claim that balancing work and home commitments affect career progression in the corporate research and development sector. But other factors such as a lack of encouragement from management also prove relevant (Wyarczyc and Renner, 2006).

These variables are the industry, type of career, cultural context effect, and the individual condition. Above all, individual factors, organizational factors, and social factors are, generally, the essential intervening "situation variables". But these all oblige women to face a choice between "stay stuck or exit". Individual factors concern the capability of achieving leadership positions in organizations, mainly in academia (Roseberry et al., 2016). But fewer women than men are allowed to express this leadership capacity. It often only happens if they are encouraged to do so. In an "extreme evaluation", quoting the authors:

≪Striving towards the ideals of womanhood, professionalism, motherhood, wifehood, scholarship, community membership, and teaching, women find themselves attempting to negotiate often conflicting identities that leave them feeling like failure is their only option≫ (*Ibidem*)

This all creates tension between individual professional aspirations and being a "caring woman" (Acker and Feuerverger, 1996). Notably, in universities, many women faculty members also feel the distance between the public and private life and social roles (Parsons and Priola, 2013).

The gender gap in the universities is, then, truly emblematic. And the differences there are more relevant, often, than in other organizations (Goastellec and Pekari, 2013). Probably, the higher demand for hard skills from men and soft skills from women also leads to a "self-fulfilling prophecy" (Misra et al., 2011). In short, individual career and personal ambitions are more easily sacrificed than role expectations. Soft skills are the ones that best match the other roles women are subjected to in the private sphere, and, to some extent, in universities. These become, then, the skills considered most accessible, and therefore those expected of women. In the meantime, they are forced to adapt to the system of expectations. In fact, both the glass ceiling effect and the leaky or clogged pipeline are present. Thus, the research question is: does the gender gap affect the careers and the timing of promotions for Italian faculty members?

There will be two steps for statistical analysis. The first is a cross-sectional analysis of the MUR (Ministry of University and Research) archive (MUR, 2020). This is done, to describe the structure of Italian faculty positions in terms of gender ratio, from 2001 to 2020. The second step for pointing out the covariates related to the academic transitions is a longitudinal Event-History Analysis (EHA).

2.3 Data and aims

This study considers Italian faculty population from 2001 to 2020, as provided by MUR. The dataset contains individual data for each year and the advancements occur at the beginning of each academic year. Each statistical unit has the following variables for each academic year:

- a unique identifier for each faculty member;
- first name and last name;
- gender;
- position: full professor (*Ful*), associate professor (*Ass*), and assistant professor (*Ast*). The *Ast* professor collects three categories: RTDa (3-year fixed-term professors, introduced in 2011), RTDb[1] (2-year fixed-term professors, introduced in 2015), and RU (tenured assistant professors, discontinued in 2011);
- in Italy, the fields of study are classified into 14 fields of study. To simplify the analysis, we considered just six fields: 1) Humanities; 2) Economics; Social and Political Sciences, and Law; 3) STM (Science, Technology, and Mathematics); 4) Engineering and Architecture; 5) Agriculture and Veterinary; 6) Medicine. The adopted classification is close to the ISCED one, with some modifications necessary to take into account the Italian context;
- the university.

The final career record is obtained by merging the annual records with the first and last name as key and other variables in cases of homonym. The observation period is given by the interval between the recruitment (or the beginning of the study, 2001) and the retirement year (or the end of the study, 2020). The maximum observation period is twenty years.

The aims of the analysis are to determine:

- What has happened to men's and women's careers advancement in the last 20 years?
- Whether it is easier for men to advance in their career from the beginning and throughout? (assistant vs associate or associate vs full)?
- Whether it is faster for men to advance in their career from the beginning and throughout? (assistant vs associate or associate vs full)?

[1] The data analysis reveals that almost all RTDa receive "promotion" to RTDb (85%). Since the path is the same for almost all, they will be treated as a single position.

- Whether men's and women's academic transitions are different among fields of study?

One limitation of our study is the lack of inclusion of data concerning national scientific qualification (*Abilitazione scientifica nazionale*, ASN[2]) to analyse, for all categories in general, how transitions have changed before and after the introduction of ANS. Furthermore, if it was possible to reconstruct the database for the year in which the qualification was obtained and the corresponding academic transition, it would be possible to examine whether obtaining the qualification earlier accelerated the career and/or whether obtaining more qualifications had any effect.

Other limitations of our study are the lack of inclusion of data concerning maternal leave, motherhood, and productivity. In the scientific literature, there exists a comprehensive summary of maternal leave, motherhood, and productivity (Marini and Meschitti, 2018). First, "family ties" has a negative effect on productivity (Fox, 2005). Howevere, according some studies "family ties" do not impact promotion in US (Perna, 2005; Sax et al., 2002) and in Scandinavia (Heijstra et al., 2015). Secondly, motherhood and family formation "have strong and independent negative effects on the likelihood that women obtain ladder—rank positions", while "these findings do not extend to later career transitions. Irrespective of marriage and children, women have less likely to get tenure and less likely to be promoted to full professors" (Wolfinger and Wilcox, 2008). About investigating productivity, in US in 1995 the "productivity measures account for a portion of the gender gap in tenure, but in each discipline a substantial share of the gender gap remains unexplained by these factors" (Weisshaar, 2017). In Italian academia between 2013 and 2016, men have 24% greater probability to be promoted to full professor, even when controlling for productivity, than women (Marini and Meschitti, 2018). Finally, differences in promotion to tenure by gender could persist after we control for productivity, demographic characteristics, and discipline (Ginther and Hayes, 1999).

On the other hand, to the best of our knowledge, the novelty and the strength of our study is the collection and the analysis of 20 years of data, which allows us to measure how gender penalizes women in the frequency of promotions and in the length of time necessary to receive a promotion and how gender penalization has changed over the last 20 years.

[2] The national scientific qualification (Abilitazione Scientifica Nazionale, ASN) is a key process designed to ensure that candidates for university professorships have the necessary academic credentials and qualifications according several criteria. The ASN certifies that an individual has reached a level of academic competence required to apply for a position as an associate or full professor in Italian academia.

2.4 Statistical analysis

As already mentioned, we, first, proceed with the cross-sectional analysis. Second, we proceed with the EHA.

2.4.1 Cross-sectional analysis

The original database is made up of individual observations. The total number of faculty members hovered around 60,000 and 65,000 over the observation period and the proportion of each group is around one-third. In Figure 2.1, the lines describe the M/F ratio for the three levels. The full professors' indexes fall more steeply over the years.

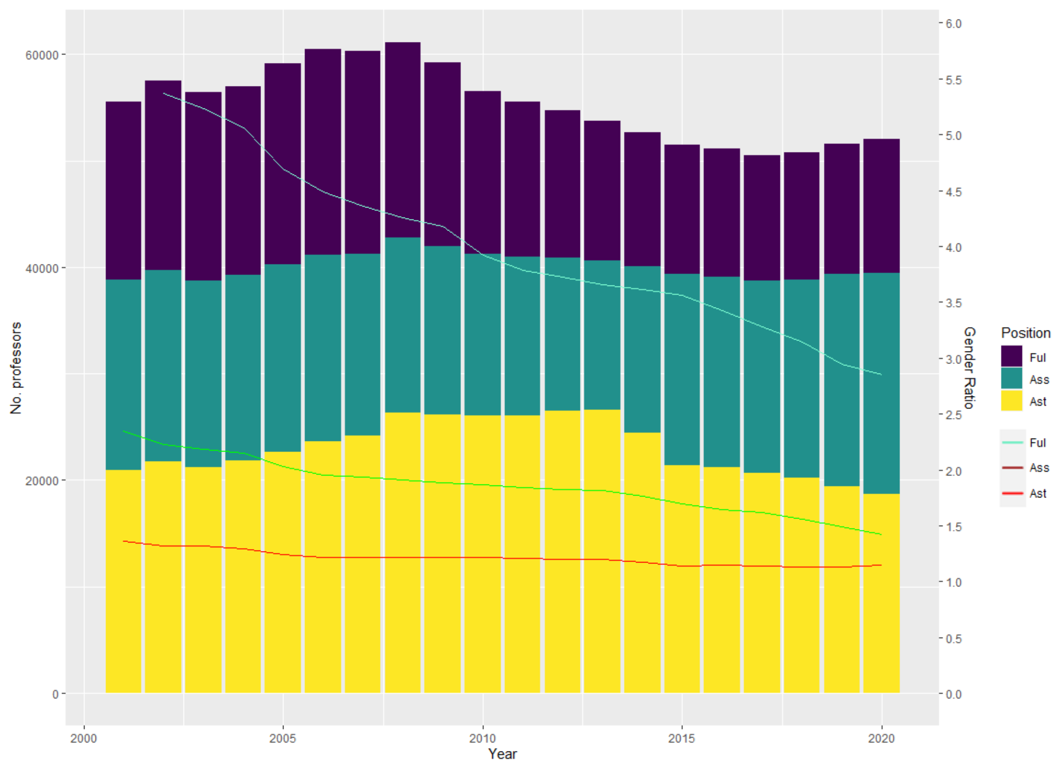

Figure 2.1: Number of faculty members (bars) and M/F ratio (lines) by position in Italy from 2001 to 2020

To measure the variation over time, we compute the variation indexes:

$$V_{M/F}\,(2001, 2020) = \frac{M/F\,(2020) - M/F\,(2001)}{M/F\,(2001)} \tag{2.1}$$

The gender ratio decrease between 2001 and 2020, measured by the variation indexes, shows that the "best improvement" is for the full professors (Table 2.1), which is equal to -0.51.

Table 2.1: Gender ratios and variation indexes by position in Italy (2001 and 2020)

	Ast	Ass	Ful	Tot
M/F (2001)	1.36	2.35	5.82	2.33
M/F (2020)	1.15	1.42	2.85	1.53
V (2001, 2020)	-0.15	-0.40	-0.51	-0.34

Table 2.2 shows the gender ratios computed for three positions and by different fields of study. As already seen, the gender ratio decreases from 2001 to 2020 for all fields and positions. "Engineering and Architecture" and "Humanities" show, respectively, maximum and minimum gender ratios in both years (and for three positions). These values mirror the well-known gender gap in these fields. Interestingly, only in Humanities the first two gender ratios are less than 1, while the gender ratio for full professors is still greater than 1.

Table 2.2: Gender ratios by position and fields of study in Italy (2001 and 2020)

Field of study	2001					2020				
	M/F				N	M/F				N
	Ast	Ass	Ful	Tot		Ast	Ass	Ful	Tot	
AgrVet[1]	1.37	2.73	9.41	2.71	2963	1.05	1.27	3.26	1.44	2974
EcSSLaw[2]	1.41	2.64	6.67	2.61	8367	1.10	1.28	2.85	1.53	9438
EngArc[3]	3.40	6.03	14.48	5.87	7851	2.25	2.74	5.51	3.00	8921
Humanities	0.66	1.04	2.38	1.11	9855	0.81	0.86	1.41	0.94	8399
Medicine	2.03	4.14	11.89	3.37	10578	1.14	1.90	3.84	1.68	7677
STM[4]	1.06	1.94	5.51	2.00	15976	1.02	1.31	2.60	1.37	14611
Overall	1.36	2.35	5.82	2.33	55590	1.15	1.42	2.85	1.53	52020

[1] AgrVet: Agriculture and Veterinary
[2] EcSSLaw: Economics, Social and Political Sciences, and Law
[3] EngArc: Engineering and Architecture
[4] STM: Science, Technology, and Mathematics

In addition, Table 2.3 shows the gender ratios computed for three positions and by macroregion. The gender ratios are almost always in favour of women in the Centre (2001) and the North (2020) for all positions.

Table 2.3: Gender ratios by position and macroregion in Italy (2001 and 2020)

Macroregion	2001					2020				
	M/F				*N*	*M/F*				*N*
	Ast	Ass	Ful	Tot		Ast	Ass	Ful	Tot	
North	1.32	2.31	6.03	2.37	22004	1.19	1.43	2.71	1.56	21854
Centre	1.32	2.24	5.59	2.24	16996	1.14	1.42	2.94	1.53	14219
South	1.46	2.51	5.80	2.38	16590	1.11	1.41	3.00	1.50	15947
Overall	1.36	2.35	5.82	2.33	55590	1.15	1.42	2.85	1.53	52020

2.4.2 Event-history analysis

Several papers have dealt with the issue of academic transitions using logistic models (Marini and Meschitti, 2018; Perna, 2001; Smith et al., 2019; Thomas et al., 2004; Durodoye et al., 2020). The limitation of using these models is the fact that they fail at taking into account of the longitudinal nature of academic transitions.

To better understand the academic transitions of faculty, by controlling for different concomitant covariates, we apply a longitudinal EHA using the Cox discrete survival time model to study the duration and the timing of events. This model provides a framework to describe the timing of event occurrences and to model the relationship between event occurrences (academic transitions) and covariates. To our knowledge, EHA has not been used in Italian academia, while there are several applications in other countries. Some papers apply continuous Cox models (Kaminski and Geisler, 2012; Box-Steffensmeier et al., 2015; Ginther and Hayes, 1999; Groeneveld et al., 2012; Long et al., 1993), while others consider discrete-time Cox models (Weisshaar, 2017; Wolfinger and Wilcox, 2008). Firstly, the discrete nature of academic transitions, which typically occur at defined intervals in time (e.g., annual promotions from assistant to associate and from associate to full professor), makes discrete models more appropriate than continuous ones. Continuous models might not accurately capture the periodic nature of such events, risking oversimplification of the underlying process. Moreover, discrete-time models offer advantages in handling time-varying covariates and addressing issues related to censoring and truncation commonly encountered in longitudinal studies. By discretizing time into intervals, these models facilitate the incorporation of time-varying predictors and the analysis of time-to-event data in a more straightforward manner. One of the limitations of this model is a large number of parameters to model the time. To reduce the number of parameters, the best approximation is usually given by a polynomial.

The two events of interest are the transitions from *Assistant professor* to *Associate professor* and from *Associate professor* to *Full professor*. To carry out a longitudinal academic transition analysis with our data, we need to know the recruitment year (categorized into four 5-year classes) and the timing of advancements. For this reason, we consider the sub-population of all the faculty members hired between 2001 and 2020 as assistant professors and still working at the end of 2020. Assistant and associate professors still working in 2020 are censored observations. Retirements and drop-outs are right-censored cases. Most of the faculty members follow the usual promotion path *Assistant professor* $\longrightarrow$ *Associate professor* $\longrightarrow$ *Full professor* and we just consider this path.

Table 2.4 summarizes the transitions proportions, conditionally to field of study and gender, from *Assistant professor* to *Associate professor* and from *Associate professor* to *Full professor*, including all assistant professors hired between 2001 and 2020. The observation period in on average equal to 6.4 years. As expected few faculty members moved from *Assistant professor* to *Full professor* (just 5434). The highest gender difference of the advancement proportions is in the Economics, Social and Political Sciences, and Law field, while the lowest gender difference is in the Agriculture and Veterinary field.

Table 2.4: academic transition proportions by gender and field of study: 2001-2020 (row percentages)

Field of study	Gender	Ast	Ass	Ful	N
AgrVet	F	82.7	16.8	0.5	7158
	M	80.8	17.8	1.4	8164
EcSSLaw	F	77.5	20.9	1.6	27889
	M	69.8	26.0	4.2	34100
EngArc	F	77.0	21.6	1.4	15104
	M	73.8	24.0	2.2	35998
Humanities	F	76.8	22.0	1.2	31740
	M	72.3	25.7	2.0	26121
Medicine	F	86.5	12.6	0.9	18579
	M	78.8	19.5	1.7	25344
STM	F	82.1	17.1	0.8	36341
	M	75.9	22.1	2.0	41154
Overall	F	80.0	18.9	1.1	136811
	M	74.4	23.3	2.3	170811

The Cox-discrete time model estimates the conditional probability that individual i will experience the target event in time period j $(T_i = j)$ given that s/he did not experience it in any earlier time period $(T_i \geq j)$:

$$h(t_{ij}) = Pr\{T_i = j \mid T_i \geq j\} \tag{2.2}$$

To estimate $h(t_{ij})$, the general model is:

$$logit(h(t_{ij})) = [\tilde{\alpha} f(t)] + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_P X_P \tag{2.3}$$

where f is a function of time, that is the time elapsed to the occurrences of the advancements ($t = 1, \ldots, 19$). X is the generic covariate considered in the estimation of the model. The usual discrete-time Cox model accommodates "time" with a number of parameters equal to the maximum length observed (in our case, the maximum number of years necessary to be promoted). To obtain a more parsimonious model, we include the term $\tilde{\alpha} f(t)$, whose better fit is a second-order polynomial in both models.

Since we consider two steps (*Assistant professor* to *Associate professor* and *Associate professor* to *Full professor*), two different Cox models are estimated (Model 1 and Model 2). Cox's models assume that the events are independent of each other, which may not be true in this context. Promotion from assistant to associate professor might influence the probability of subsequent promotion to full professor. Furthermore, although missing information that could have introduced bias into the results has been taken into account, it may still be unrealistic to assume that relative risks remain constant over time for all academic transitions. Alternatively, a multi-state model might be more elegant and appropriate for modelling sequential academic transitions in future research (Andersen and Keiding, 2002). This type of model takes into account dependencies between different phases and allows transitions between states to be modelled directly, rather than treating them as separate events.

In any case, In Model 1 the probability of being associate at time t, assumed to be assistant at time $t-1$, is estimated. The Model 1 is fitted with a second-degree polynomial for *Time*, *Gender*, *Macroregions*, *Field of study*, and *Year of Recruitment* as main effects. Interactions between *Time* and *Gender*, *Time* and *Field of study* and between *Gender* and *Field of study* are included. In Model 2, we estimate the probability of being full at time t, assumed to be associate at time $t-1$. The covariates are the same as Model 1 without *Year of Recruitment* variable and the interaction between *Gender and Field of study*, plus an extra variable *Time Ast to Ass*, which corresponds to the number of years necessary for the previous advancement. The baseline profile is given for a man, working in a northern university, in Agriculture and Veterinary field, hired between 2001 and 2005.

$$
\begin{aligned}
log\left(\frac{Ass_i(t)}{Ast_i(t-1)}\right) = {} & \alpha_1\, t + \alpha_2\, t^2 + \beta_1\, Gender_i + \beta_2\, Macroregion_i + \\
& + \beta_3\, FieldOfStudy_i + \beta_4\, YOfRecr_i + \beta_5\, Gender_i \cdot t + \\
& + \beta_6\, FieldOfStudy_i \cdot t + \beta_7\, Gender_i \cdot FieldOfStudy_i + \\
& + \beta_8\, Gender_i \cdot YOfRecr_i
\end{aligned}
\tag{2.4}
$$

$$
\begin{aligned}
log\left(\frac{Ful_i(t)}{Ass_i(t-1)}\right) = {} & \alpha'_1\, t + \alpha'_2\, t^2 + \beta'_1\, Gender_i + \beta'_2\, Macroregion_i + \\
& + \beta'_3\, FieldOfStudy_i + \beta'_4\, TimeAstToAss_i + \\
& + \beta'_5\, FieldOfStudy_i \cdot t + \beta'_6\, Gender_i \cdot t
\end{aligned}
\tag{2.5}
$$

In Table 2.5, the estimated coefficients are reported. The estimates of the parameters in Models 1 and 2 have the same sign. So, the two-academic transitions have the same pattern with some differences in magnitude. The probability of becoming an associate professor for the baseline profile at t=7 is 0.035, while for the corresponding female is 0.025. In Model 2, the time required to move from associate professor to full professor has, as expected, a positive linear effect, so for instance, the probability of becoming a full professor, conditioning on "Time Ast to Ass" = 7, for the baseline profile at t=7 is equal to 0.107, while for the corresponding female is 0.071. Based on the estimates from the second model, it can be observed that professors who spent less time in the transition from researcher to associate tend to spend less time in the second transition (coefficient "Time Ast to Ass"). Looking at the gender gap, it is "less difficult" for women to advance from assistant to associate than from associate to full. The probability of becoming an associate professor for an "average" assistant professor within seven years is 0.06 for women and 0.09 for men. The probability of becoming a full professor for an "average" associate professor within seven years is 0.182 for women and 0.302 for men. The gender effect is strong in both models, but it is attenuated by the interaction term "Time*Gender", which has a slight negative effect. This implies how, over time, the difference in career progress between males and females tends to decrease. Some dampening is given by the interaction of the field of study with time. The parameters of these interactions are almost all negative. The macroregional estimates reflect the usual divide by which academic transitions are faster in the North. The interaction between gender and field of study was also reported in the first passage. What is observed is that STM disciplines and medicine are the fields of study where the gender difference is most remarkable.

Another interesting result regards the *Year of Recruitment*, occurring only in Model 1, in fact estimates show, for males and females, that it is easier to obtain an advancement especially in the last cohorts. Moreover females show an extra "gain" given by the positive interactions *YearOfRecr*Gender*. On the other hand, in the transition from associate to full professor, the *Year of Recruitment* is not significative.

Table 2.5: Parameter estimates of the Cox discrete-time models

Covariates		Model 1			Model 2		
		Estimate		Std. Error	Estimate		Std. Error
	Intercept	-7.436	***	0.141	-0.776	***	0.363
	Time	0.776	***	0.017	0.924	***	0.055
	Time^2	-0.027	***	0.001	-0.036	***	0.002
Gender (M)	F	-0.679	***	0.116	-0.844	***	0.128
Field of study	EcSSLaw	1.079	***	0.133	0.927	**	0.341
(AgrVet)	EngArc	0.181		0.136	0.365		0.352
	Humanities	0.622	***	0.134	0.032		0.355
	Medicine	0.544	***	0.141	0.858	*	0.362
	STM	0.442	***	0.132	0.487		0.348
Macroregion	Centre	-0.173	***	0.022	-0.080		0.059
(North)	South	-0.380	***	0.020	-0.189	***	0.055
Year of recr.	(2005;2010]	0.545	***	0.027	\	\	\
((2000;2005])	(2010;2015]	1.801	***	0.038	\	\	\
	(2015;2020]	2.948	***	0.063	\	\	\
Time * Gender (M)	F	0.047	***	0.007	0.057	***	0.017
Time * Field of study	EcSSLaw	-0.068	***	0.013	-0.089	*	0.044
(AgrVet)	EngArc	0.024		0.013	-0.027		0.046
	Humanities	-0.016		0.013	-0.013		0.045
	Medicine	-0.054	***	0.013	-0.116	*	0.047
	STM	-0.012		0.012	-0.050		0.045
Gender * FieldOfStudy	F*EcSSLaw	-0.212	*	0.092	\	\	\
(M) * (AgrVet)	F*EngArc	-0.036		0.096	\	\	\
	F*Humanities	-0.109		0.092	\	\	\
	F*Medicine	-0.371	***	0.099	\	\	\
	F*STM	-0.281	**	0.091	\	\	\
Year of recr. * Gender	F*(2005;2010]	0.042		0.042	\	\	\
((2000;2005]) * (M)	F*(2010;2015]	0.285	***	0.061	\	\	\
	F*(2015;2020]	0.635	***	0.097	\	\	\
Time Ast to Ass		\	\	\	0.134	***	0.009

Baseline categories are in brackets. Time is computed in years.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The last output of the EHA is the estimated survival curves, that in our case, represent the probability of changing position in the observation period. This tool allows us to better describe academic transitions quantitatively. The gender gap is apparent in all curves (Figure 2.2). The comparison of survival curves, for both models, shows that the second advancement is faster than the first. The gender gap favours males, and it becomes slight higher over time (Figure 2.2).

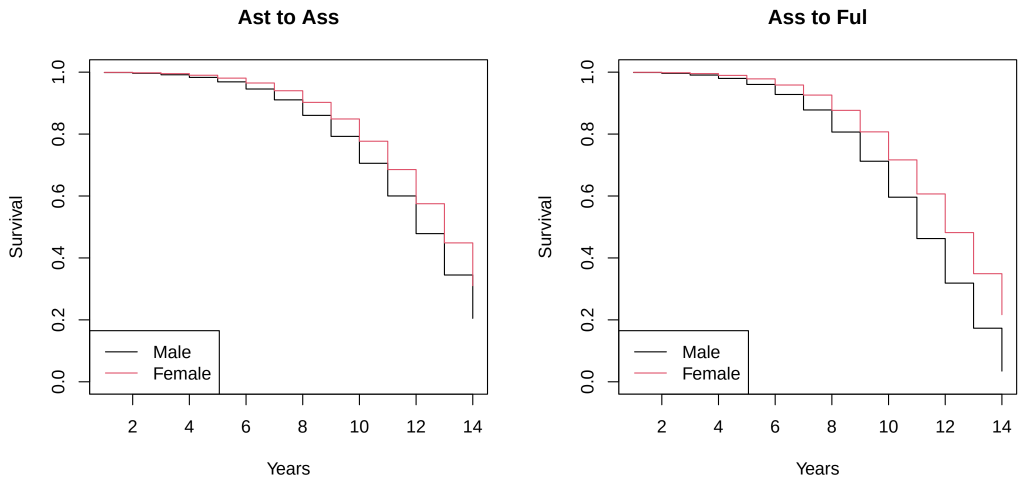

Figure 2.2: Estimated discrete-time survival curve by gender and type of advancement for the baseline profile (northern university, Agriculture and Veterinary field, hired between 2001 and 2005).

For instance, after nine years and six months, 25% (1-75%) of the men *Assistant professor* have "moved" to *Associate professor*, while the time needed for women is ten years and three months. In the second case, the time needed to advance increases the gender gap (six years and ten months for the men, seven years and ten months for the women) (Table 2.6). The smaller difference between men and women, which seemed to reduce the stickiness in the gender ratio in enrolment and faculty recruitment, becomes again significant here. The same assessments apply in the second advancement.

Table 2.6: Estimated percentiles in years and months to academic transitions by gender and type of advancement for the baseline profile (northern university, Agriculture and Veterinary field, hired between 2001 and 2005).

	Ast $\longrightarrow$ *Ass*		*Ass* $\longrightarrow$ *Ful*	
	Men	Women	Men	Women
75%	9 yrs 6 mo	10 yrs 3 mo	6 yrs 10 mo	7 yrs 10 mo
50%	11 yrs 10 mo	12 yrs 7 mo	8 yrs 5 mo	9 yrs 6 mo
25%	13 yrs 8 mo	14 yrs 5 mo	9 yrs 8 mo	10 yrs 8 mo

Table 2.6 summarizes the years needed by 25%, 50% and 75% of faculty members to make one of the two advancements based on calculations made on the estimated model, considering the baseline profiles (reported on Figure (2.2)). It can be seen

from the table that women, in all scenarios, take longer to advance their careers than men: the differences are higher from *Associate professor* to *Full professor*. 50% of the male assistants became associate professors in eight years and five months, while the same percentage of females became assistant professors one year and one month later. For the first step, however, the same percentage of women take nine months longer than men. However, it is notable that the baseline profile refers to the first cohorts, so Table 2.6 does not take into account the "gain" obtained by females recruited in the last years.

2.5 Discussion and conclusions

Despite the regulatory choices made and some undeniable progress, Italy remains a country with a high gender gap. This disparity is also to be found in universities. In many fields of study, there is an important difference between men and women, particularly for the top position. Not only do a greater proportion of men reach top positions, but they also do so faster. The Italian social structure is unquestionably the explanation for these differences. The disparity of access to higher positions is still culturally accepted and highlights a difference in the social role and the symbolic value assumed by those positions. In short, the distinction between "masculine" and "feminine" jobs still matters in the Italian imagination and much less than in other Nordic countries where occupational segregation is higher (Steinmetz, 2011). Though it has gradually decreased over time. It is evident that a broader base of women's participation in university courses in some scientific fields of study does not correspond to a greater presence of women employees in universities' said fields. For instance, in mathematics, Italian freshmen have been, for forty years, equal to Italian "freshwomen". But the M/F ratio for full professors is more than four to one.

Our data shows a sort of "friction effect", "wall resistance" in all the fields of study. It is as if women constituted a "social fluid" that flows less easily in the hydraulic system constituted by the structure of the Italian university. This fluid loses its natural thrust towards the summit, running into a narrow mesh filter at the end of the path. The parallel with the "glass ceiling" is obvious.

In the non-technical and non-scientific fields, as (Bourdieu, 2001) would note, the course of study is characterized by the fact that it is very close to the traditional definition of women activity in society. It is less frequented by men for just this reason. In the second case, a different effect can be hypothesized. In the choices regarding professional paths, for women who study business, choosing an academic career is preferable to other careers that their degree would allow. In fact, in 2020 the assistant professors M/F ratio was equal to 1.15 showing a "recent women preference" in the universities. This might be taken to suggest that a career as a company executive or entrepreneur is more demanding - in terms of time, willingness to travel, the ability to take risks, flexibility in changing jobs, coping with stress, etc. In the Italian imagination, all those conditions are considered much more manageable for men than for women. Women do not reach the professions that are better paid and more linked to power: power understood here as the power of disposition, leadership, etc. They do not do so because of the self-selection process to which (Bourdieu, 2001) refers. They are, therefore, more likely to choose a university career. Here they find greater opportunities for the same reasons because these are spaces that are not preferred by their male colleagues and they can manage family responsibilities, too. In short, a gendered form of distinction continues to be reproduced, in which, in the collective imagination, the best position is masculine. What is left is "feminine".

The stickiness comes back if we look at the gender gap in academic transition regarding access to the top of the career ladder and transition speed. There is surely a trend reducing the gender gap, but it looks slow. We do not actually know when the gender ratio for assistant professors will be the same for full professors.

Moreover, two ratios are computed within each gender: Ass/Ast and Ful/Ass, in 2001, 2010 and 2020 (Gaiaschi and Musumeci, 2020). These ratios give a different perspective of the academic transitions within each gender. Table 2.7 shows the presence of the leaky pipeline (all the male ratios are always greater than the female ones) and the persistence of a glass ceiling effect measured by the female Ful/Ass ratios that are always between 0.35 and 0.60.

Table 2.7: Ratios Ass/Ast and Ful/Ass by year and gender

	2001		2010		2020	
	Men	Women	Men	Women	Men	Women
Ass/Ast	1.035	0.599	0.693	0.450	1.218	0.982
Ful/Ass	1.148	0.463	1.227	0.585	0.764	0.381

In conclusion, both the "leaky pipeline" and the "glass ceiling" effects seem to intervene powerfully in the gender gap of Italian universities in both careers' transitions. In the first step, there is an evident glass ceiling effect, in fact the distance between the two curves is present till the end (Figure 2.2, *Assistant professor* $\longrightarrow$ *Associate professor*), with a significative improvement for females in the last cohorts. On the other hand, in the second step, after 15 years, the distance between women and men is less evident (Figure 2.2, *Associate professor* $\longrightarrow$ *Full professor*). This is probably due to a "seniority effect", which is dramatically present in the Italian public administration labour market.

Chapter 3

Balancing procedures for analysing the high school-university transition in Italy

This chapter is based on the work from Falco, V., Vittorietti, M., Attanasio, M. (2024). High school-university transition in Italy: a balancing approach. *Work in progress*

Abstract

Educational inequalities begin in early childhood and persist throughout students' careers, which are characterised by a series of steps ("transitions"). These transitions are influenced by several factors. Specifically, this chapter explores the transition from high school to university in Italy, analysing a merged database from the 2019/2020 academic year, focusing on how socio-economic status, high school track, and geographical region affect university transition rates. To address the non-random selection of students in the process of attending high school, we compare two balancing techniques: the multilevel propensity score method and a meta-analytical method, to ensure that socio-economic status is balanced across different high school tracks in different geographical macroregions. The "balanced" results show a higher socio-economic level is associated with a higher probability of being enrolled in university and a complex joint effect between the macroregion and the high school track.

3.1 Introduction

A high rate of graduation should ensure a country's economic development by promoting job creation, raising the overall skill level of the workforce, stimulating innovation and maintaining global competitiveness. For individuals, obtaining a degree could be essential to secure good job prospects, develop professional skills and competences, achieve personal growth through specialised knowledge, reduce socio-economic inequalities and advance socially.

According to the European Commission's 2021 education statistics, Italy has fewer graduates than other European countries. Student enrolment has declined significantly, especially after the economic crisis of 2008, with a steady recovery in the last five to six years. Inadequate financial support, reduced public funding for the university system and higher tuition fees are possible causes of the decline in enrolment rates at Italian universities (De Angelis et al., 2016).

On an individual level, obtaining a degree is a complex path that varies from person to person. Educational inequalities begin to manifest themselves in early childhood and continue to affect students throughout the entire school career. There are differences in performance at every level of schooling, to the point where pupils are unaware of the external circumstances that unconsciously influence their decisions. These include the macro-economic scenario, such as economic deprivation and employment conditions in the labour market, and individual demographic characteristics, such as ethnicity, socio-economic status, geographical origin and gender, which more or less favour further schooling in the educational pathway. Because of these effects on access to university education, educational inequalities have long-term consequences for individuals and society (OECD, 2018; Pillas et al., 2014). Although there are still some socio-economic differences that may influence students' decisions to continue or drop out of education, the OECD notes that even after accounting for the increased supply of high school graduates, the demand for university education increased in many European countries in 2018-19 compared to previous years (OECD, 2019a).

In this context, the high school–university transition is crucial. This phenomenon is influenced by several factors. Factors, such as geographical distance from the university, income level, urban-rural divide, and gender, are sometimes tangible and measurable variables that can impact students' decision to pursue higher education. For example, students from rural areas or low-income families might face more significant barriers to accessing university education due to financial constraints or limited educational resources (Morton et al., 2018). Other factors such as discouragement, lack of confidence and family influence may also be important, although they may be difficult to measure. These psychological and emotional barriers cannot be quantified in this context, but they can have a significant impact on students' motivation and confidence in their ability to succeed in higher education (Lister et al., 2023).

The aim of this chapter is to analyse the transition from high school to university. It is well known that students' performance at university depends on a complex network of multidimensional factors, including the student's high school career, socio-economic 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., 2008; Agasisti and Vittadini, 2012). However, very little research has been carried out in Italy to examine how these factors affect academic outcomes outside of surveys or small-scale analyses focused on a single institution. This is most likely due to the lack of longitudinal microdata. However, we have access to a big dataset that combines school and university careers at high school level (details will be provided in Section 3.3). This allows us to perform the analysis with the highest level of detail of information we have access to. We use a merged database that allows us to gain unique insights into how some factors, such as socio-economic status, high school track and geographical macroregion, influence the transition rate at high school level in Italy in 2019/2020.

Students' choice of high school is not random, and this choice may vary according to geographical macroregion and socio-economic status, highlighting a hierarchical structure in the data. This structure recognises that students are nested within different tracks and schools are further nested within geographical macroregions. This hierarchical data structure is essential to take into account in order to accurately analyse educational outcomes and understand the influences on student performance. If the "groups" are very different, statistical analyses can produce biased results. In our case, the groups are high school tracks. If high school tracks are considered as clinical trials, the problem of group unbalance (in terms of socio-economic status) can be addressed in several ways. In particular, a comparison will be made between balancing techniques to solve this problem and to investigate the factors that influence the school-university transition. We compare multilevel propensity score method (Arpino and Mealli, 2011) and with a meta-analytical method based on the non-parametric comparison of empirical cumulative distribution functions (Aiello et al., 2011). The multilevel propensity score method takes into account the hierarchical structure of the data, ensuring that socio-economic and geographical factors are balanced across different school tracks. The meta-analytical method, on the other hand, provides a robust alternative by directly comparing the distribution of variables between groups. After applying the balancing methods, we will consider a modelling approach to compare investigate the effect of the aforementioned factors on the transition. This approach will compare the data obtained from the balancing methods, allowing us to examine the impact of socio-economic and geographical factors on the school-university transition in Italy in greater detail.

The chapter is organised as follows. Section 3.2 provides an economic and sociological background. Section 3.3 presents the data used and preliminary analysis. Sections 3.4 and 3.5 present some theoretical insights and the results of the two balancing techniques, while Section 3.6 shows the results of the modelling approach.

Finally, a brief summary and some conclusions are given in Section 3.7.

3.2 Theoretical background

When we talk about educational growth, we usually refer to increases in enrolment. Over the last few decades, Western countries have been characterised by increasing participation rates in higher education (Schofer and Meyer, 2005). This upward trend has been confirmed at all levels of education in Italy (Ballarino and Panichella, 2014), where it appears that new enrolments come mainly from lower socio-economic groups (Ballarino and Panichella, 2016). This growth can be explained by the increasing demand for educational qualifications (Collins, 1979) in the labour market and as a result of changing economic conditions. However, there are still distinct patterns across different strata of the population, i.e. some segments of the population show persistent differences in educational behaviour and choices. The economic framework of rational choice (Boudon, 1974, 1979) and the sociological machinery of the cultural capital hypothesis (Bourdieu, 1966; Bourdieu and Passeron, 1964, 1977) are two opposing views that coexist in the academic literature on the decision to continue or discontinue education.

3.2.1 Economic framework

It is important to recognise that investment in education can have financial benefits (Stanfield, 2011). Each additional year of education can be seen as a proxy for a worker's skills and hence potential earnings. Therefore, if we conceptualise education as a variable amount of human capital, then each additional year of education will have a direct impact on the rate of return to education (Bourdieu and Passeron, 1977). Students should be encouraged to pursue higher education if the predicted financial benefits of their decision to attend university justify it. In light of this, some scholars argue that the desire to further one's education is primarily motivated by the hope of earning more money in the future (Becker, 1975, 1992; Heckman, 2000). In other words, the higher the expected monetary return from an additional qualification, the more likely a student is to continue their education (e.g. make the transition from high school to university). The value of these qualifications fluctuates over time, i.e. they lose value and also suffer from inflation. As more people with the same qualifications apply for available jobs, the value of these skills decreases over time as a result of increased competition (Contini et al., 2017). One immediate effect is a weakening of the employment prospects for high school graduates as participation in high school and university increases. Another viewpoint that allows us to examine educational choices treats education as a product (Alstadæ ter, 2009). Education can be seen as a consumption good rather than an investment good, with the consumption value being represented by the non-monetary immediate benefits associated with being a student (the advantage of socializing and making new friends, taking part in schooling events and activities, visiting new places, learning new things, etc.). Although this result is significantly biased toward high-status students, the individual consumption value of education is a non-negligible factor that students genuinely consider (Belfield et al., 2016).

There are several explanations for why students move from high school to university. The main economic theory in this context is the "rational choice model" (Breen and Goldthorpe, 1997). This model suggests that decisions about schooling are made after careful consideration of three key elements. The first is the cost of continuing education. These costs can be direct or indirect. Regarding direct costs, it should be noted that although recent changes in education systems have greatly increased the likelihood of lower class individuals continuing their education (Ballarino et al., 2016), it would be oversimplifying to assume that income no longer affects the likelihood of continuing education. Indirect costs, also known as opportunity costs, are essential for all the lost wages that students could have earned by engaging in some other activity instead of attending school for a second year. The probability of success is the second component of the rational choice model, while the advantages of more education form the third component. In effect, people have different options which they choose based on the predicted future benefits and the utility they attach to each (Boudon, 1979). This component supports the theory that human capital and the desire to avoid social mobility should be the primary motivators of educational choice (Boudon, 1974). According to this theory, people from different social classes experience different costs, all other things being equal. In particular, working-class households suffer far less status loss from dropping out of school than middle-class households. This hypothesis provides factual evidence for the strategies used by the wealthier strata of society to prevent intergenerational mobility that would otherwise raise the (educational) bar even higher. It also explains why differences in educational attainment persist despite the reforms implemented to increase educational participation and the significant wave of educational growth in recent decades.

3.2.2 Sociological framework

One of the main factors influencing inequalities in educational attainment is social background. Social background primarily affects academic performance and shapes personal choices based on that performance (Jackson, 2013). Numerous hypotheses have been proposed to understand how social background might affect academic performance (Bourdieu, 1966, 1977; Bourdieu and Passeron, 1964; Tyler, 1981). They argue that social origin affects children both through the genetic transmission of traits and through the home environment. The transmission and replication of cultural capital (Bourdieu and Passeron, 1977), cultural and social resources and their impact on achievement gaps (Bourdieu and Passeron, 1964) are examples of the mechanisms suggested to be particularly important. Children's motivation to study and achieve good grades is particularly enhanced by their daily contact with a dynamic home environment made up of educated parents and characterised by ongoing cultural incentives. As a result, family cultural resources and years of education establish a form of cultural reproduction privilege towards their children, motivating them to follow in their parents' footsteps (Bourdieu, 1966). On the other hand, the lack of cultural resources in working-class families prevents pupils from progressing in their studies or acquiring further titles. Another case is that of all the students who say from an early age that they would go to university without assessing the pros and cons (Breen and Goldthorpe, 1997). Some students automatically assume that they will go to university, displaying a so-called "college-going habitus", rather than making a choice (Grodsky and Rieglecrumb, 2010). Similar findings are consistent across many Western countries, including Italy (Contini et al., 2018; Argentin and Triventi, 2011). In the sociological literature it is argued that tracks in high school affect the likelihood of students enrolling in university. When students are assigned to a distinct, non-overlapping track at each stage of their academic career, a school system is said to be horizontally stratified. The more horizontally stratified an education system is, the more an individual's social and parental background influences his or her decision to pursue further education. The vast majority of school systems in European countries divide students into specific tracks that allow them to specialise in pre-determined fields, while many other subjects are neglected. In Italy (or in Europe in general) there is no selection in the educational process and each student is free to enrol in the type of school he or she wants. In the absence of legally enforceable constraints, family background - which includes parental occupation, previous schooling and the general home environment - is the primary element that determines the sorting of pupils into different school tracks (Ballarino et al., 2016; Checchi and Flabbi, 2005) and their performance (Brunello and Checchi, 2007). Once certain groups of students are assigned to particular tracks, several potentially harmful mechanisms can come into play. For instance, the production of human capital is positively affected when more brilliant students are grouped in the same class and work side by side, but when lower-level students are grouped together, their ability cannot find a way to take off. Teacher sorting, or the tendency for better teachers when have the power to decide to teach in high-ability classrooms, is a direct result of this segmentation. Adjusting for school track at the upper secondary level does not help to disentangle the effects of social origin from the effects of horizontal stratification and their possible interaction (Jackson, 2013). Pupils seem to find it more difficult to change tracks as the tracks become more distinct. Indeed, the choice of school is a decision that will affect students for the rest of their lives, especially those from lower socio-economic backgrounds (Gambetta, 1987).

3.3 Data and preliminary analysis

3.3.1 Data

In this chapter three different sources of data are considered:

- The micro-level longitudinal data from the National Archive of University Students (**ANS**). ANS is a comprehensive database that contains detailed information about the university careers of all students enrolled in Italian universities. Data are available from 2008 to 2022. This database is designed to provide a deep understanding of the educational trajectories and outcomes of individual students over time. The database includes a record for each enrolled student. Within the dataset, there are variables about each student (gender, etc.). ANS has limited knowledge of high school background (just final high school mark and high school mark are available). Furthermore, the dataset contains data about the university career as the degree course, the academic performance, changes of course.

- The micro-data from the National Evaluation Institute for the School System (**INVALSI**). INVALSI is a valuable resource for educational research and analysis in Italy because it collects detailed information about high school students' profiles who go on to university. Data are available from 2018 to 2021. This encompass a wide range of individual characteristics with respect to ANS, including information about student's socio-economic background and family, details about the geographical origin, and the high school performance (indicators of a student's high school performance, including whether they had a regular high school career). The dataset includes standardized test scores in Italian, Mathematics, English reading and listening.

- The aggregate data from **"Scuole in Chiaro"**. Scuole in Chiaro is an online platform where users can access various types of data and information about schools such as the school track (scientific, technical, vocational, etc.), geographical information, the number of students who got their diploma by gender and final mark, and their continuation rate by school track, i.e., the proportion of students that go to university. Data are available from 2014/15 to 2018/19.

For each school, individual students' data are available for those enrolled at the university, while only aggregate data are available for those not enrolled at the university. For this reason, the analyses will be conducted only an aggregate dataset where the statistical unit is the high school track and the goal is to investigate the factors that influence the transition rate (Figure 3.1).

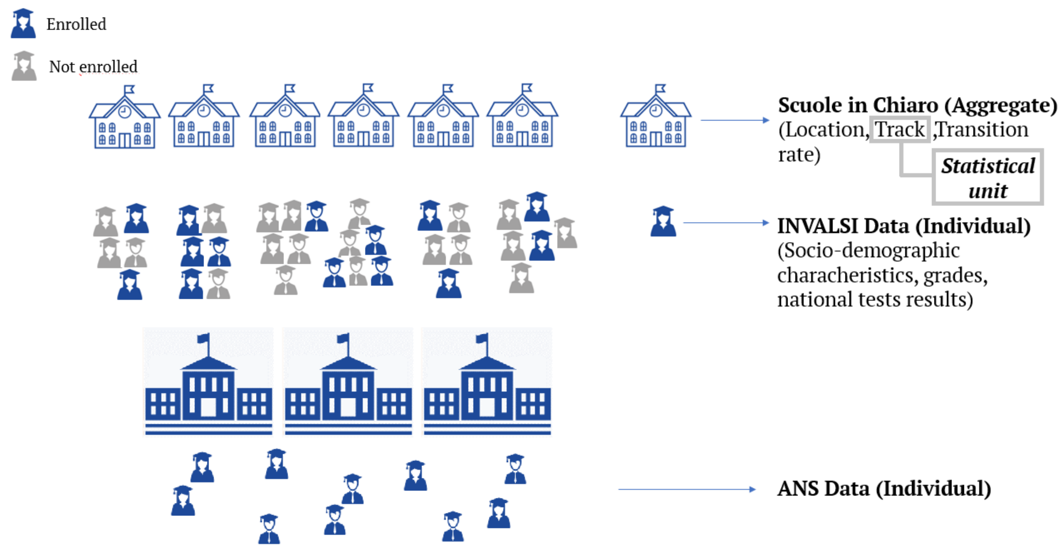

Figure 3.1: Data structure

The linkage of the aforementioned databases (ANS-INVALSI-Scuole in Chiaro) allows investigation of the transition from high school to university. 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 data of the 2020/21 cohort.

The statistical unit is the high school track. We have not considered the whole school as a statistical unit because there are more and more high schools that have various tracks within them. As a result, the high schools within them are extremely heterogeneous in terms of transition rates among tracks.

We have excluded all the high schools misclassified for which the transition rate is missing (0.6% of the data) and the private schools because they are structurally different from public high schools and represent 1% of total high schools (in terms of students, 0.7%).

Missing data imputation

The socio-economic status ("*Escs*[1]") of high school tracks contains missing data (5.8%). There is no reason to assume that the missingness mechanism is dependent on the variable of interest in this case (Escs). Furthermore, looking at the distribution of missing data on socio-economic status in relation to the response variable (transition rate), it appears that the percentage of missing data is similar at different levels of the transition rate (Table 3.1).

Table 3.1: Available and missing data for socio-economic status of high school tracks by transition rate

Transition rate	Available data		Missing data		Total	
	n	%	n	%	n	%
<30	1980	94.2	121	5.8	2101	100.0
30-80	3239	94.1	202	5.9	3441	100.0
>80	2198	94.4	131	5.6	2329	100.0
Total	7417	94.2	454	5.8	7871	100.0

The missingness mechanism appears to be independent of the observed data (also with respect to macroregion and type of high school). This is equivalent to saying that the behaviour of two units that share the same observed values have the same statistical behaviour on the other observations, whether observed or unobserved. So, it is reasonable to assume at least that the generating mechanism of the missing data is missing at random (MAR). We apply a double imputation missing data procedure. First, we applied a hot deck imputation (Ono and Miller, 1969). Hot deck imputation is a method where missing values are imputed using observed values from similar records (concerning variables other than socio-economic status such as transition rate, macroregion, high school track, high school "higher" mark percentages) in the dataset. Secondly, we apply a deterministic regression imputation. It involves predicting the missing values based on the relationship between the variable with missing data (Escs) and other variables in the dataset based on the results (predicted values) of a regression model.

[1] The indicator Escs is computed using principal component analysis of three indicators: HISEI (employment status of parents), PARED (educational level of parents) and HOMEPOS (possession of certain specific materials). It is an indicator with zero mean and unit standard deviation. For example, a student with a strictly positive Escs value is a student with a more favourable socio-economic-cultural background than the Italian average (OECD, 2019b).

3.3.2 Preliminary analysis

This section aims to provide a preliminary analysis of the transition rate to university on the entire high school tracks population in Italy in 2019/20 (7,827 high school tracks) using "Scuole in Chiaro" dataset. Transition rates of each high school track will be weighted by the number of students[2]. First, we report the distribution of the weighted transition rates. The weighted transition rate distribution of the high school tracks in Italy in 2019/20 is negatively skewed. The average weighted transition rate is equal to 56.1 (Figure 3.2).

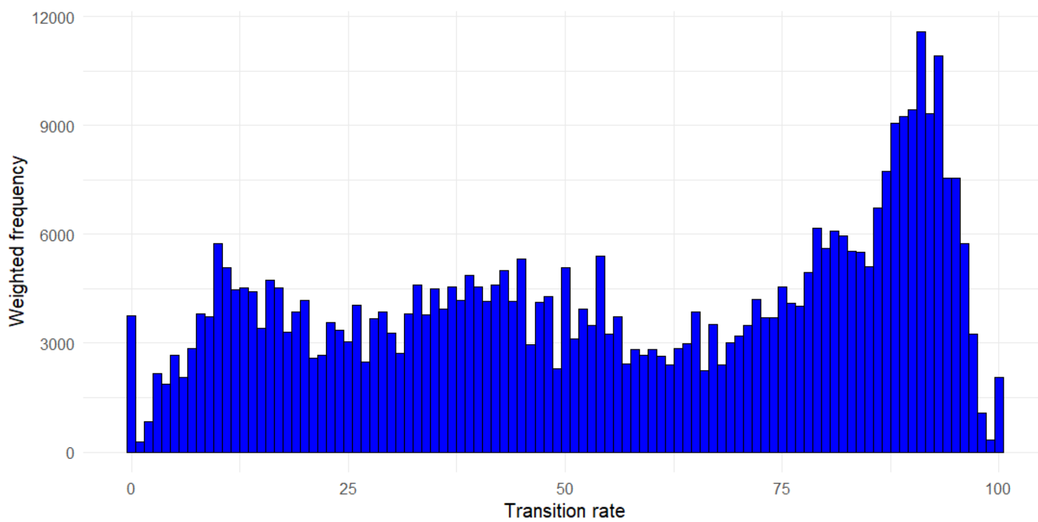

Figure 3.2: Distribution of weighted transition rates of high school tracks in Italy in 2019/20

If we analyse the transition rates in Italy by region (Table 3.2 and Figure 3.3), we can note that the highest transition rates are in the Center, while the lowest are in the South and Islands except for Abruzzo and Molise. In particular, Campania and Sicilia are characterized by the lowest rates (53.0 and 53.1 respectively), while Umbria share the highest ones (62.3). It is important to note that the provinces of Aosta and Bolzano are not included since data from those regions are not available. There are two limitations related to the available data. The transition rate is extremely heterogeneous within each region at the provincial level and this pattern is not captured in this analysis. The other limitation is that the data of Italian students who attended high school abroad are not available; this could lead to an nderestimation of the transition rates especially for the North of Italy.

[2] The weighted transition rate Tr_w can be calculated as follows: $Tr_w = \frac{\sum_{i=1}^{N} w_i Tr_i}{\sum_{i=1}^{N} w_i}$ where Tr_i is the transition rate for the i-th high school track, w_i is the i-th weight, N is the total number of high school tracks.

Table 3.2: Weighted transition rates of high schools in Italy in 2019/20 by region

Macroregion	Region	n	%	Tr_w	Sd_w
North	Emilia-Romagna	454	5.8	56.7	29.0
	Friuli-Venezia Giulia	166	2.1	56.6	28.0
	Trentino	76	1.0	58.4	24.8
	Veneto	567	7.2	53.5	28.6
	Liguria	179	2.3	59.5	28.5
	Lombardy	958	12.2	56.2	28.2
	Piedmont	523	6.7	56.8	28.9
	Valle d'Aosta	\	\	\	\
Centre	Latium	627	8.0	61.0	27.8
	Marches	236	3.0	61.0	29.2
	Tuscany	507	6.5	57.2	28.8
	Umbria	139	1.8	62.3	28.8
South and Islands	Abruzzo	204	2.6	60.3	29.6
	Apulia	627	8.0	53.8	30.4
	Basilicata	142	1.8	57.5	31.0
	Calabria	359	4.6	55.7	29.8
	Campania	939	12.0	53.0	31.9
	Molise	65	0.8	60.0	29.7
	Sardinia	280	3.6	52.5	26.6
	Sicily	779	10.0	53.1	30.3
Total		7827	100.0	56.1	29.4

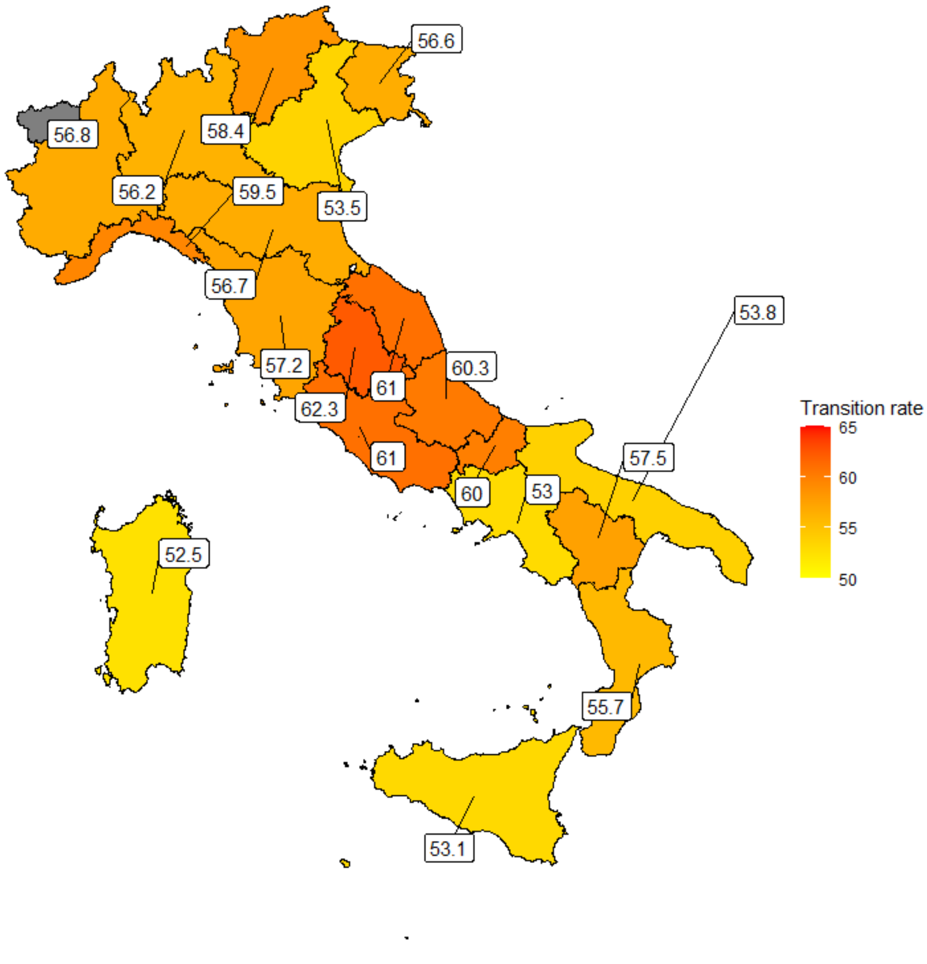

Figure 3.3: Weighted transition rates of high school tracks in Italy in 2019/20 by region

As mentioned in Section 3.2, the Italian education system is made of different high school tracks. Here, “Humanities” corresponds to the “Liceo Classico”, “Traditional Science” to the “Liceo Scientifico Tradizionale”, “Applied Science” to the “Liceo Scientifico delle Scienze Applicate”, and “Vocational” to “Istituti Professionali”. The category “Other liceo and Technical” includes other types of lyceums (e.g. language, art, music, etc.) as well as “Istituti Tecnici”, both technological and economic addresses. These two types of high school tracks have been grouped together since both other types of lyceums and technical institutes have similar transition rates.

Looking at Table 3.3 it is self-evident that the transition rate is higher for students who got their diploma from Humanities, where the transition rate is 89.2%, even larger than the one of Traditional Science (88.8%). Transition rates for students who got their diploma from Applied sciences is slightly lower (84.4%) than for Humanities and Traditional science, while the transition rate for students who got their diploma from Other liceo and Technical have a greatly lower transition rate (52.3%). Lastly, the type of schools characterized by the lowest transition rate are vocational high schools where the percentage of students who continue their educational path is around 15.2%.

Table 3.3: Weighted transition rates in Italy in 2019/20 by high school track

	n	%	Tr_w	Sd_w
Humanities	533	6.8	89.2	7.8
Traditional Science	1114	14.2	88.8	8.3
Applied Science	676	8.6	84.4	11.8
Other liceo and Technical	4076	24.8	52.3	20.3
Vocational	1428	18.2	15.2	9.6
Total	7827	100.0	56.5	13.5

In Humanities, Umbria and Abruzzo stand out with remarkably high transition rates (94.2% and 94.0% respectively). On the other hand, Sardinia has a lower rate (82.5%). The transition rates from Traditional Science high schools present a fairly consistent picture across the regions, with most hovering around the 90% mark. However, there are still differences, with the highest transition rate in Liguria (92.0%) and Sardinia (82.7%). Applied Science high schools show a more pronounced variability in transition rates. Emilia-Romagna and Tuscany have higher rates (90.4% and 90.1% respectively), while Campania and Puglia have lower rates (79.2% and 80.7% respectively). The transition rates from Other Liceo and Technical institutes also show similar regional differences. Lastly, the lowest transition rates are found in Vocational high schools. Trentino and Liguria show relatively higher transition rates than Sicily and Campania. This suggests that while the core curriculum may be uniform, regional factors still play an important role. We also note that the percentage differences between the highest and lowest transition rates are not equal for each high school track. The largest differences are observed in Applied Science and Other Liceo and Technical. In Humanities and Traditional Science, the percentage difference between regions is less pronounced than in the other categories, suggesting that the disparity between regions is less pronounced in these high school tracks (Figure 3.4).

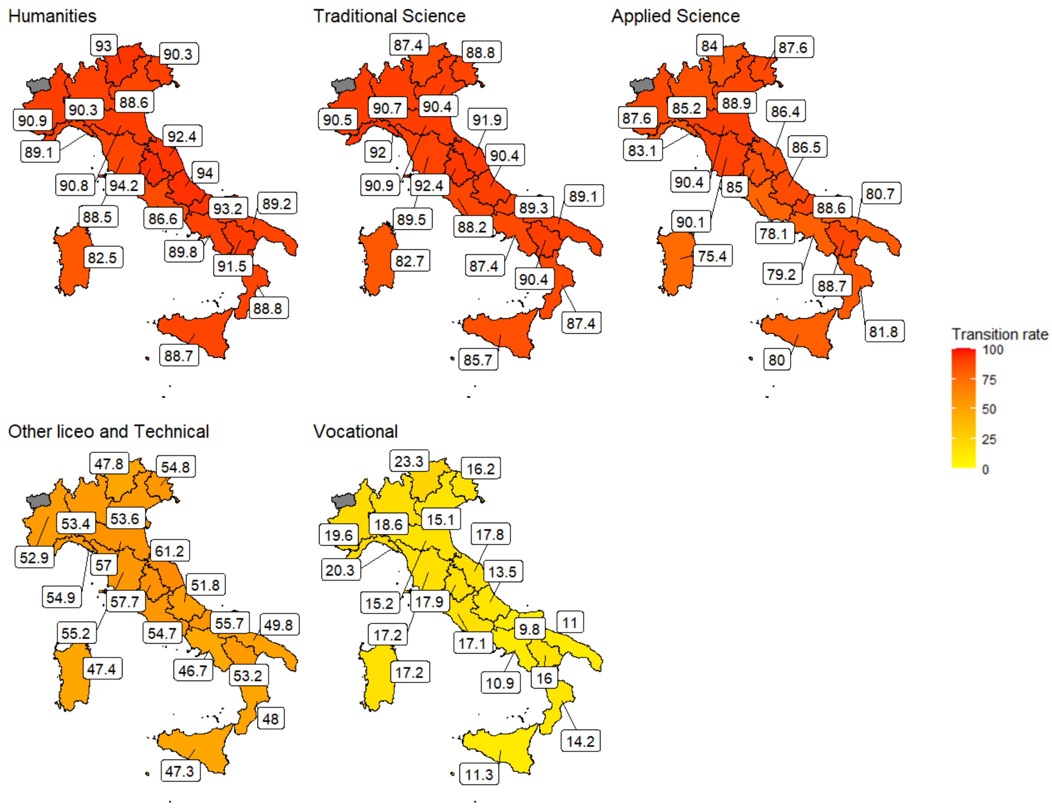

Figure 3.4: Weighted ransition rates in Italy in 2019/20 by region and high school track

We also can highlight the difference in socio-economic status with respect to the high school track (Figure 3.5). Students from upper-class families tend to enrol in humanistic or scientific high schools, which better prepare them for university education (Contini and Scagni, 2012). In contrast, students from lower-class families are more likely to enrol in technical or vocational schools. This pattern reflects the hierarchical structure of Italian high schools. However, this educational disparity based on socio-economic status might perpetuate social inequalities and limit opportunities for students from disadvantaged backgrounds (Ballarino and Panichella, 2016).

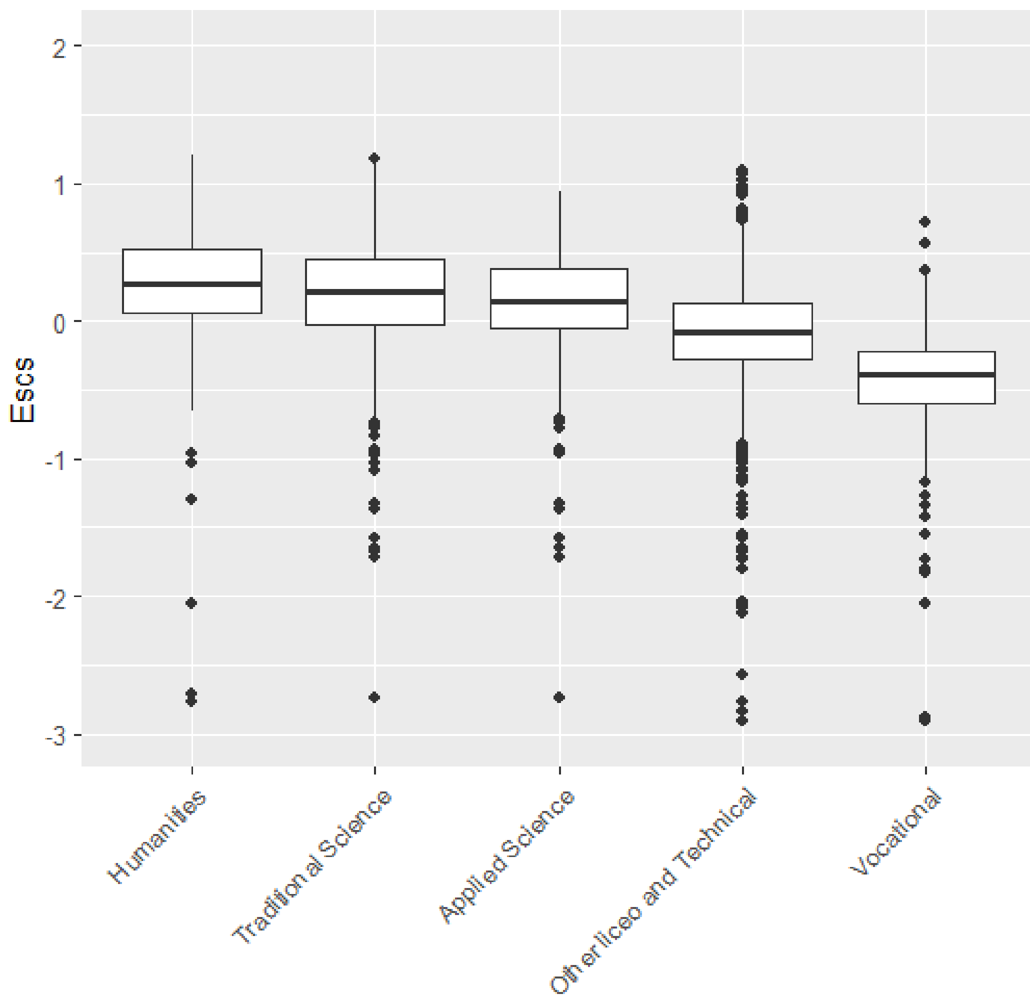

Figure 3.5: Boxplots of socio-economic status (Escs) in Italy in 2019/20 by high school track

Finally, we analyse the socio-economic status for each high school track in Italy by region (Figure 3.6). The highest values of Escs are in the North (0.3 in Trentino), while the lowest in the South and Islands (-0.7 in Abruzzo).

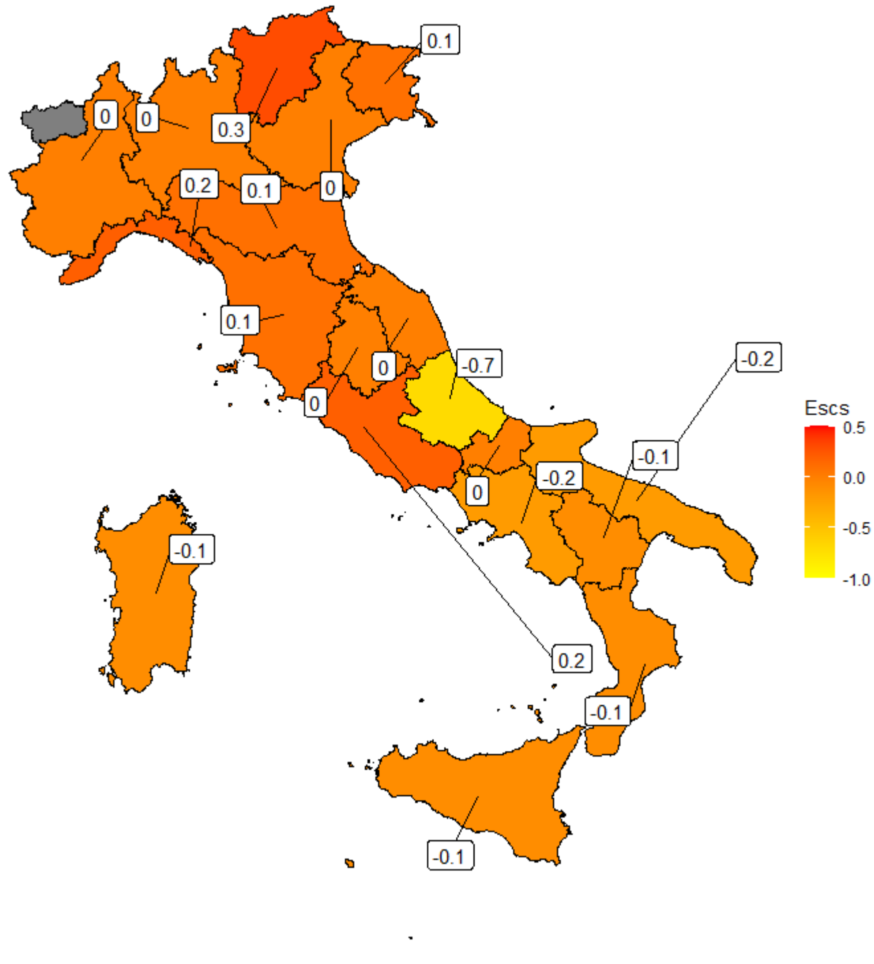

Figure 3.6: High school track's socio-economic status (Escs) in Italy in 2019/20 by region

3.4 Statistical methods

As reported in the introduction, the nature of the data does not allow the use of usual methods based on microdata. As reported in the Section 3.3, individual data are available for students enrolled in the university, while only aggregate data are available for those not enrolled in the university. We obtain a merged aggregate database where the statistical unit is the high school track.

3.4.1 Meta-analytical method

The meta-analytical method deals with selection bias in order to understand whether the allocation to groups is random. Focusing on the aggregate transition data, the selection of students into high schools is not random and this causes selection bias. If high school tracks are considered as clinical studies, the problem of group unbalance can be addressed in a theoretical framework of meta-analysis. When dealing with non-random selection of students into high schools (analogous to a treatment group in a clinical study), addressing the issue of unbalance in your dataset becomes crucial. Unbalance in this context refers to the fact that the students are not randomly assigned to high schools, which can lead to potential biases in your analysis. In particular, there are two sources of non-randomness:

- **within each school**: the transition to university depends on Escs
- **between schools**: the choice of the type of high school is strongly affected by the Escs (as seen in the preliminary analysis)

Meta-analysis, widely used in various fields including the social sciences and originated in the field of education and psychology (Glass and Smith, 1978), is an analytical technique used to combine multiple studies, typically clinical trials. Combinability in meta-analysis refers to the "similarity" between separate studies. Specifically, combinability can be defined as "the extent to which separate studies can be assumed to measure approximately the same thing" (Macarthur et al., 1995). In other words, it assesses whether the studies are similar enough in terms of their characteristics, methodologies, and populations to allow for meaningful statistical synthesis. In a meta-analysis medical framework, qualitative issues (e.g. intervention or exposure variability, outcome measures, time frame and follow-up, publication bias, adherence to protocols, etc.) of combinability are always examined extensively (Haidich, 2010), while quantitative ones are essentially limited to sample sizes and effect sizes. Actually, the quantitative assessment of clinical combinability studies is unsatisfactory (Feinstein, 1995; Borenstein et al., 2009), because it lacks specific quantitative criteria to establish when trials can be considered similar enough.

The meta-analytical method based on the non-parametric comparison of empirical cumulative distribution functions (ECDFs) identifies studies responsible for the lack of combinability due to unbalances between treatment groups regarding specific covariates by comparing their distributions in the treatment groups, without making any assumptions about their shapes (based on the findings of a previous study in which the authors created a method to evaluate the covariate unbalance with regard to a single covariate (Aiello et al., 2011)). Meta-analysis unbalances may stem from cumulative smaller unbalances rather than a single study (Trowman et al., 2007). A “rule of thumb” procedure aims to eliminate these unbalanced studies, enhancing the reliability of meta-analytical results.

In our case, we classify the potential factors of unbalance as:

- $1st$ level covariate ($1st$-LV): *Macroregion*
- $2nd$ level covariates ($2nd$-LVs): *High school track*, *Escs* (Socio-economic status)

In the dataset, we refer to the $1st$-LVsbecause it reflect different “populations” from a structural point of view, while the $2nd$-LVs are the “risk factors” related to the event under study (enrolment at university).

The method is based on the Empirical Cumulative Distribution Function (ECDF) of Escs built for each sub-group (Formula 3.1):

$$F_j(\theta) = \frac{\#\left(\hat{\theta}_{(j)} \leq \theta\right)}{N_j} = \begin{cases} F_j(\theta) = 0 & \theta < \hat{\theta}_{(1)} \quad j = 1 \\ F_j(\theta) = \frac{N_j}{N_J} & \hat{\theta}_{(1)} < \theta < \hat{\theta}_{(j+1)} \quad j = 1, \ldots, J-1 \\ F_j(\theta) = 1 & \hat{\theta}_{(J)} < \theta \end{cases} \tag{3.1}$$

where j are the sub-groups ($j = 1, \ldots, J$) with $n_{(1)}, n_{(2)}, \ldots, n_{(J)}$ subjects, $\hat{\theta}_{(j)}$ are the ordered statistic for the j-th sub-group, F_j is the ECDF, $N_J = \sum_{j=1}^{J} n_{(j)}$ is the total number of subjects in the sub-group and N_j is the cumulative number of subjects in all sub-groups with a statistic $\hat{\theta}_{(j)}$ less than or equal to a given θ. Hence, the ratio $\frac{N_j}{N_J}$ is the value of the ECDF at $\hat{\theta}_{(j)}$. If all the sub-groups have distinct $\hat{\theta}_{(j)}$, then the number of jumps of $F_j(\theta)$ will be equal to J; otherwise, if there are two or more sub-groups that share the same $\hat{\theta}_{(j)}$, then the number of jumps of $F_j(\theta)$ will be equal to the number of distinct $\hat{\theta}_{(j)}$.

The covariate balance holds if the ECDFs are not statistically different. To compare distribution functions the null hypothesis to be tested is

$$H_0: \; F_k(x) = \bar{F}(x) \quad \forall x \in (-\infty, +\infty) \tag{3.2}$$

where k is the generic category of a group variable and $\bar{F}$ is the overall ECDF, which aggregates the distributions of all categories. In other words, the k sub-groups have the same distribution function in some way (Zhang and Yuehua, 2007).

Non-parametric tests based on variations in the cumulative distribution functions are generally seen to be highly useful for testing this hypothesis. Among these, the Kolmogorov-Smirnov test (K-S test) stands out as one of the most prevalent. The K-S test is non-parametric tool used to assess whether two samples are drawn from the same unspecified continuous distribution. It is renowned for its sensitivity to differences in the tails of the distribution and its ability to detect even small discrepancies, particularly with large samples. The greater the value of statistic, the larger the discrepancy between the distributions of the two samples. Also, there exists a weighted version of the K-S test to adjuste the maximum absolute difference between the empirical distribution functions by their respective weights. The weighted K-S test is defined as

$$K - S_{N,m,n} = \max\left(|w_m F_m(x) - w_n G_n(x)|\right) \tag{3.3}$$

where $F_m(x)$ and $G_n(x)$ are the empirical distribution functions of the two samples, w_m and w_n are the respective weights for $F_m(x)$ and $G_n(x)$, and $N = m + n$.

The method's justification is to evaluate combinability, pinpoint the studies that are unbalance-causing, and then eliminate or modify the impact of that unbalance on the metanalysis' outcome variable.

The procedure here follows two sequential steps:

- Assessing the **global combinability**: the global combinability holds when the randomization process holds with respect to the levels of macroregion, that is, the Escs' ECDFs controlling for the macroregion's levels are not statistically different. This is investigated both graphically and analytically, through the Kolmogorov-Smirnov test.

- Assessing the **local combinability**: the local combinability holds if the randomisation process holds with respect to the levels of high school track given a category of macroregion. Also in this case, the local combinability is investigated both graphically and analytically, through the Kolmogorov-Smirnov test.

In both steps, the goal is to "identify" and "exclude" the unbalanced trials. Specifically, in the "identifying" algorithm we propose two subset selection criteria: "*t first*" proposed in the literature and "*t min*" proposed in this thesis.

- **Identifying** the unbalanced trials: all the studies responsible for the highest observed unbalance are identified, with an iterative procedure, which ends only when it reaches a statistically reasonable balance among the sub-groups. Once identified, the unbalanced trials are excluded. The steps of the iterative procedure are (given a value of α):

  - Build I sets $\{S1 \setminus S1_i\}$, for $i = 1, 2, \ldots, I$, whose cardinality is $\mid I - 1 \mid$

- Compute the quantities $T_{h(-i)}$, for $h = 1, 2, 3$, and $\{T_{c(-i)}\}$, $\forall i = 1, \ldots, I$ (where T is the value of the Kolmogorov-Smirnov test)
- Identify the $\min_i\{T_{c(-i)}\}$, and then the corresponding i-th study
- Remove the i-th study, $S1_i$, and consider the set $S1_{(-i)} = \{S1 \setminus S1_i\}$
- Rename $S1 = \{S1 \setminus S1_{(-i)}\}$
- **Subset selection criteria**
  - The first criterion (**t first**) is proposed in the literature. Formally: If $\min_i\{T_{c(-i)}\}$ is not significant, then stop, otherwise go to the first step (Aiello et al., 2011)
  - The second criterion (**t min**) is a possible methodological evolution of the method proposed in this thesis. All observations are removed and the iteration with the minimum test value is selected.

- **Excluding** the unbalanced trials: it is necessary to keep in mind the following two general criteria that are not strictly statistical evaluations: qualitative and quantitative issues. Regarding the qualitative issues, if one eliminates the studies that represent a specific sub-group, the eliminated studies could denature the meta-analysis. Instead, regarding the quantitative issues, one must pay attention to the trade-off among two aspects: the total number of eliminated studies (high school tracks) and the number of subjects (students) that correspond to the eliminated studies.

3.4.2 Multilevel propensity score

In general, the non-random assignment to groups highlights a possible hierarchical structure of the data. Specifically, educational data represents a perfect example of a hierarchical structure in which individuals (students) are grouped in clusters (high school tracks) in which they receive a "clustered treatment", namely their high school track (Pimentel et al., 2018). Here, the high school track "summarizes" the general context regarding the socio-economic level and the neighbourhood in which the schools are located ("social microcosm"). The selection of the high school track in this chapter is a treatment assignment that is not made at random because it is influenced by a variety of social and cultural elements. The decision is heavily influenced by a number of factors, including socio-economic status (Checchi and Flabbi, 2007; Ballarino and Panichella, 2014; Giancola and Salmieri, 2020; Panichella and Triventi, 2014), the geographical area in which they live (Bratti et al., 2008), the gender of the children (Contini et al., 2017), and other factors. Therefore, the treatment groups present substantial differences. Differences between the treatment and control groups can be viewed as causal effects when interventions are given at random. However, distinct results may not always reflect treatment effects, but rather beginning disparities between the treated and control groups when participants choose their own treatment (Cochran and Rubin, 1973). Estimation of a causal effect from observed data is identified by assuming unconfoundedness (no unmeasured confounders), claiming that the treatment is effectively randomized with respect to observed covariates. The estimation of casual effect is weighted for both groups to a common distribution of covariates, namely the global distribution of X in the combined population (Li et al., 2013). The propensity score is the conditional chance of a unit being assigned to a treatment given a specific collection of covariates (Rosenbaum and Rubin, 1983). This idea, which is expressed as follows: $e_i = Pr\left(Z_i = 1|X_i\right)$ (Austin, 2011), refers to the probability that a given individual will be assigned to a specific treatment group based on their pre-treatment characteristics. Z_i is the treatment assigned to the i-th observation, and X_i is the observed baseline covariates (Rosenbaum and Rubin, 1983). This score functions as a balancing metric, and while it is typically not possible to determine the true propensity score in the context of observational studies, it can be approximated using the study data that is available (Austin, 2011; Markoulidakis et al., 2023). A propensity score estimate can be found using a variety of methods. In the absence of controlled experimental settings, these strategies are useful for establishing comparable treatment groups. Propensity score techniques are employed to reduce bias in the estimate of causal treatment effects associated with these confounders by moderating the differences in observed confounding variables among treatment groups (Austin, 2011; Markoulidakis et al., 2023). In this chapter, a propensity score is used for multiple treatments since there are more treatment groups given by the different high school tracks. Moreover, given the multivelel structure of the data under analysis, multilevel propensity score (MPS) is a natural extension of the propensity score (Arpino and Mealli, 2011).

The steps are:

- **Estimate Propensity Scores**: calculate propensity scores for each individual in the dataset, considering their observed covariates (e.g. regression models, classification and regression tree analysis, etc.) and recognising the hierarchical structure of the data. This step ensures that the estimated propensity scores consider variability within and between clusters.

- **Match or Weight the Data**: Use the estimated propensity scores to either match treated and control subjects within clusters or apply weighting methods. Matching or weighting helps to create a balanced comparison group, considering the multilevel nature of the data.

The proposed propensity score method is based on generating balancing weights. This approach aims to adjust the unbalances observed between treatment groups by applying weights that increases similarity between the groups. Among the different methods to estimates weights the techniques used here are: the entropy balancing ("*ebal*" method), a measure of dispersion, and the propensity score weighting using generalized linear models. The first one it is primarily based on deriving weights through the minimization of negative entropy, subject to exact moment balancing constraints. The process of entropy balancing involves the formulation of an optimization problem, the solution to which provides the necessary weights. Whereas, the second one relies on estimating propensity scores with a parametric generalized linear model and then converting those propensity scores into weights using a formula that depends on the desired estimate (Noah, 2023).

The MPS procedure has been carried out using the R package MatchIt (Ho et al., 2011). For further theoretical insights into the estimation and evaluation of balancing weights in observational studies refer to (Austin, 2011; Austin and Stuart, 2015; Chan et al., 2016; Hainmueller, 2012; Reifeis and Hudgens, 2020; Robins et al., 2000; Thoemmes and Ong, 2016).

We propose two propensity score weighting procedures:

- **Combined procedure**: the "treatment" variable is a combination of macroregion and high school track. This means that the treatment group is defined on the basis of both the macroregion and the type of diploma at the same time.

- **Two-step procedure**: this procedure involves a two-step approach that follows the same rationale as the previous balancing procedure based on the comparison of ECDFs. In the first step, the "treatment" variable is the macroregion alone, followed by a subsequent analysis in which the treatment variable is the high school track, conditionally on the macroregion in the second step.

3.5 Results

The need to obtain (aggregate) data on students enrolled and not enrolled in each pathway by merging the three datasets led to the loss of some tracks, even though the preliminary analysis considered the entire population of high school tracks. This was likely caused by the failure to update school mechanographic codes in the ANS and INVALSI datasets. Now, the dataset consists of 6,763 high school tracks. They involve 398,077 students.

3.5.1 Meta-analytical method

First, we build the Escs' ECDFs controlling for the (three) levels of the macroregion and then we compare the ECDFs by the Kolgomorov-Smirnov test (Table 3.4).

Table 3.4: Empirical cumulative distribution functions (ECDF) by 2nd-LV macroregion (global combinability)

	Macroregion	Mean Escs	n_j (students)	ECDF
1	North	-1,963	24	0,000
2	North	-1,315	4	0,000
3	North	-1,311	72	0,001
...	...	...	...	...
2494	North	1,617	15	0,999
2495	North	1,656	1	0,999
2496	North	1,812	2	1,000
...	...	...	...	...
...	...	...	...	...
1	South and Islands	-1,998	8	0,000
2	South and Islands	1,909	7	0,000
3	South and Islands	1,884	13	0,000
...	...	...	...	...
2661	South and Islands	1,451	10	0,999
2662	South and Islands	1,532	11	0,999
2663	South and Islands	1,656	4	1,000

Figure 3.9 illustrates that the ECDFs are rather different from each other. The Kolmogorov-Smirnov tests support the hypothesis that each ECDF is “structurally” and statistically different from the average ECDF. Therefore, Escs is not balanced.

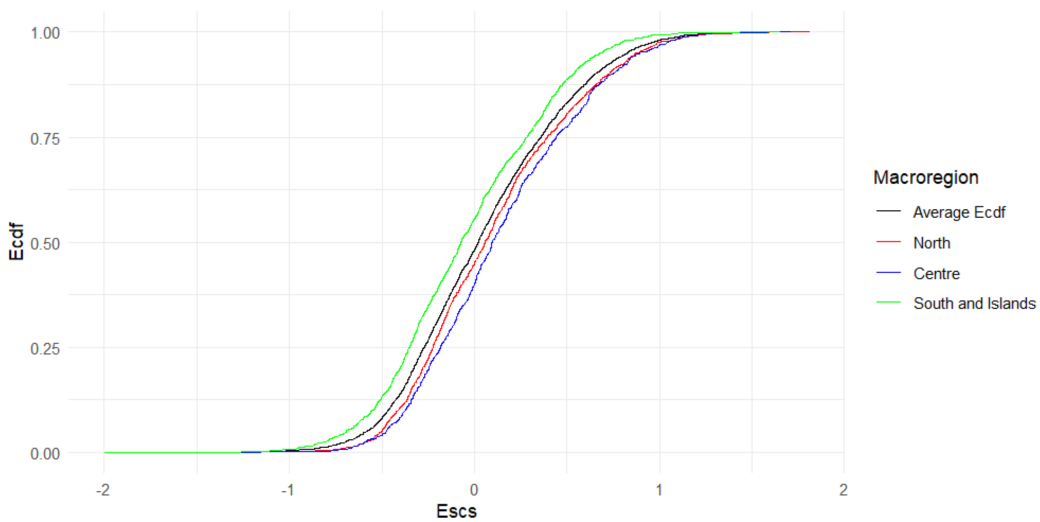

Figure 3.7: ECDFs by macroregion (1st-LV)

As mentioned above, the two identifying and excluding unbalanced trials procedures were applied, which differ in the *t first* and *t min* subset selection criteria.

Particularly in the second procedure, it is interesting to note the trend of the K-S test as the number of observations decreases (as the number of iterations increases) for the three categories of the macroregion (3.8). Initially, the trend of the K-S statistic decreases progressively with a modest slope until it reaches the minimum. The initial decrease is progressive as the test runs on datasets with few observations removed. The minimum is the point at which all observations have been removed to have the maximum statistical proximity between a sub-group curve and the overall curve. After this point the trende becomes increasing and explodes in the last part of the domain since the low number of observations makes the ECDF unstable.

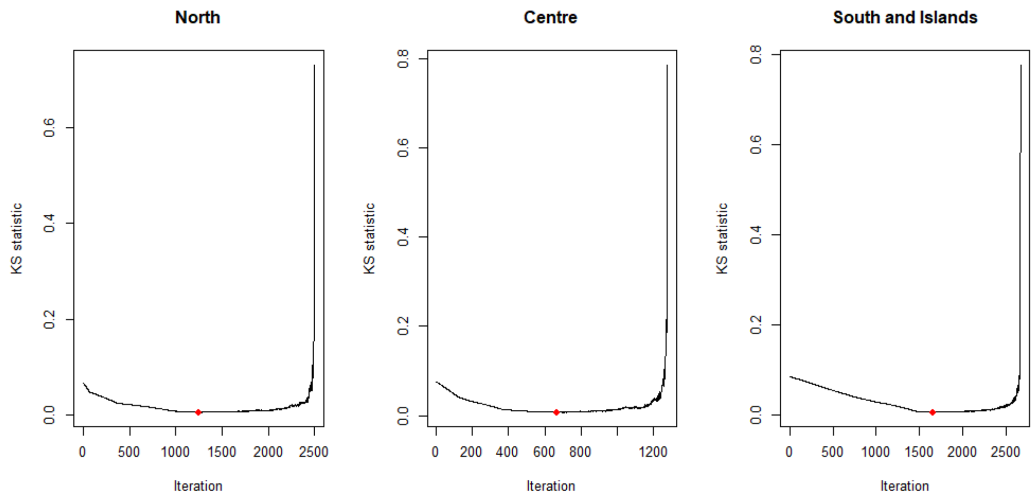

Figure 3.8: Trend of K-S Statistic by macroregion across iterations of the t min procedure

Table 3.5 reports the results of global combinability by macroregion and procedure. The best criterion in terms of distance between the ECDFs is the *t min* procedure, although the decrease in observations is larger (-55.44% of observations) than the *t first* procedure (-19.25% of observations).

Table 3.5: Results of global combinability (ex ante and ex post procedure)

		Ex ante procedure			Ex post procedure		
Macroregion	Subset selection criterion	n	%	K-S test	n	%	K-S test
North	*t first*	2496	38,8	0,067	2225	42,8	0,033
	t min				1245	43,4	0,006
Centre	*t first*	1279	19,9	0,075	1166	22,4	0,043
	t min				613	21,4	0,007
South and Islands	*t first*	2666	41,4	0,085	1810	34,8	0,036
	t min				1012	35,3	0,006
Total	*t first*	6441			5201		
	t min				2870		

An example of the graphical output of the global combinability for the "North" macroregion is now given. The figures for the other two categories are given in Appendix A. We note a most evident graphic similarity between the average ECDF (the initial overall ECDF) and ex post procedure ECDF obtained through the *t min* subset selection criterion. Although the largest part shows a greater distance than *t min* procedure and a smaller distance of *t first* procedure from the average ECDF, the preferred procedure based on the results of the Table 3.5 is *t min* procedure.

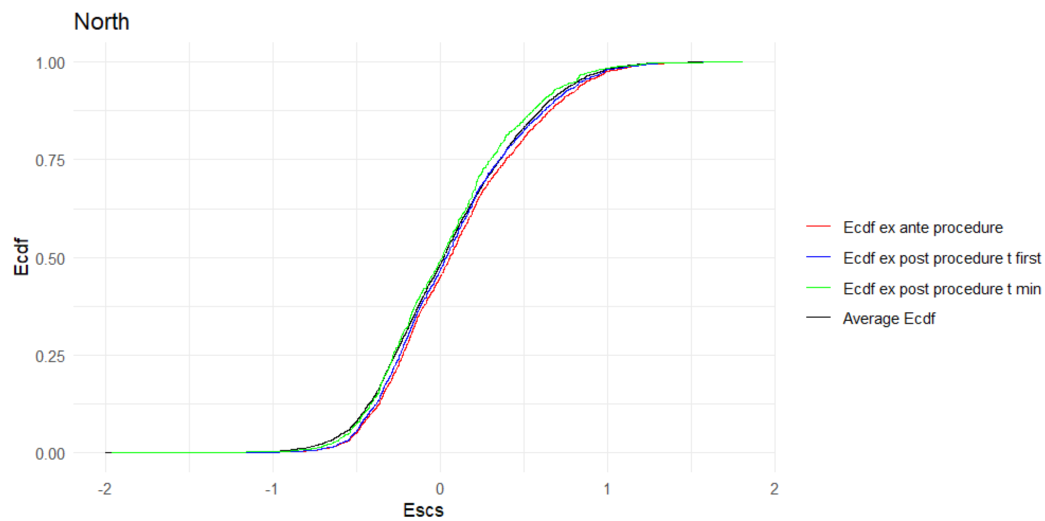

Figure 3.9: ECDFs ex ante and ex post procedures for the North macroregion

Now, we build the ECDFs by controlling for the levels of the high school track in each sub-group (where the sub-groups are the levels of macroregion). Then, the local combinability is investigated, comparing all ECDFs in each sub-group graphically and analytically by the Kolgomorov-Smirnov test following the same line of reasoning of global combinability. The local combinability results Table is shown:

Table 3.6: Results of local combinability (ex ante, for both global combinability subset selection criteria, and ex post procedure)

		Ex ante *t first*			Ex post *t first*			Ex post *t min*		
		n	%	K-S test	n	%	K-S test	n	%	K-S test
North	Humanities	169	7,6	0,571	22	2,9	0,287	8	3,1	0,096
	Traditional Science	395	17,8	0,413	59	7,8	0,177	21	8,0	0,058
	Applied Science	282	12,7	0,318	80	10,6	0,154	28	10,7	0,067
	Other liceo and Technical	988	44,4	0,181	543	72,1	0,065	193	73,9	0,016
	Vocational	391	17,6	0,498	49	6,5	0,193	11	4,2	0,104
	Total	2225			753			261		
Centre	Humanities	129	11,1	0,533	26	7,9	0,260	9	8,2	0,087
	Traditional Science	236	20,2	0,428	30	9,1	0,246	15	13,6	0,188
	Applied Science	109	9,3	0,346	43	13,1	0,209	12	10,9	0,088
	Other liceo and Technical	511	43,8	0,244	212	64,6	0,101	66	60,0	0,045
	Vocational	181	15,5	0,580	17	5,2	0,329	8	7,3	0,231
	Total	1166			328			110		
South and Islands	Humanities	238	13,1	0,487	38	7,2	0,220	19	8,8	0,109
	Traditional Science	367	20,3	0,371	83	15,8	0,152	24	11,2	0,051
	Applied Science	171	9,4	0,311	65	12,4	0,170	27	12,6	0,082
	Other liceo and Technical	768	42,4	0,238	287	54,6	0,086	123	57,2	0,021
	Vocational	266	14,7	0,488	53	10,1	0,187	22	10,2	0,068
	Total	1810			526			215		
Total	Humanities	536	10,3		86	5,4		36	6,1	
	Traditional Science	998	19,2		172	10,7		60	10,2	
	Applied Science	562	10,8		188	11,7		67	11,4	
	Other liceo and Technical	2267	43,6		1042	64,8		382	65,2	
	Vocational	838	16,1		119	7,4		41	7,0	
	Total	5201			1607			586		
		Ex ante *t min*			Ex post *t first*			Ex post *t min*		
		n	%	K-S test	n	%	K-S test	n	%	K-S test
North	Humanities	84	6,7	0,572	19	3,8	0,303	7	4,6	0,217
	Traditional Science	200	16,1	0,435	30	6,0	0,244	10	6,5	0,088
	Applied Science	163	13,1	0,360	50	10,0	0,195	10	6,5	0,092
	Other liceo and Technical	558	44,8	0,167	360	72,3	0,080	118	77,1	0,025
	Vocational	240	19,3	0,510	39	7,8	0,220	8	5,2	0,135
	Total	1245			498			153		
Centre	Humanities	34	5,5	0,540	12	5,0	0,370	3	3,7	0,204
	Traditional Science	125	20,4	0,502	8	3,3	0,440	31	37,8	0,440
	Applied Science	44	7,2	0,405	20	8,3	0,296	8	9,8	0,163
	Other liceo and Technical	289	47,1	0,171	186	77,2	0,113	34	41,5	0,047
	Vocational	121	19,7	0,565	15	6,2	0,333	6	7,3	0,192
	Total	613			241			82		
South and Islands	Humanities	147	14,5	0,497	31	8,8	0,240	12	8,5	0,090
	Traditional Science	201	19,9	0,370	47	13,4	0,202	21	14,9	0,130
	Applied Science	92	9,1	0,303	46	13,1	0,203	13	9,2	0,114
	Other liceo and Technical	421	41,6	0,253	187	53,3	0,108	77	54,6	0,023
	Vocational	151	14,9	0,510	40	11,4	0,215	18	12,8	0,080
	Total	1012			351			141		
Total	Humanities	265	9,2		62	5,7		22	5,9	
	Traditional Science	526	18,3		85	7,8		62	16,5	
	Applied Science	299	10,4		116	10,6		31	8,2	
	Other liceo and Technical	1268	44,2		733	67,2		229	60,9	
	Vocational	512	17,8		94	8,6		32	8,5	
	Total	2870			1090			376		

Although the global combinability scenario has already been chosen (t *min*), the results of all local combinability scenarios have been reported in Table 3.6. From the results of the K-S tests it seems that the t *min* scenario is preferable once again. However a compromise is necessary that takes into account a number not too small of observations. So the procedure chosen in this second phase (local combinability) of the balancing procedure is the *t first*. The overall percentage variation is -83.08% of high school tracks. The results in the form of graphs of the ex post local combinability procedure are given in the Appendix A just for the chosen ex post global combinability procedure (*t min*).

3.5.2 Multilevel propensity score

In the combined procedure, the first way to estimate weights is based on multinomial logit model. With the use of this first method, however, the balance of the variable Escs was not achieved as it can be seen in Figure 3.10, since it has a standardized mean difference greater than 0.05.

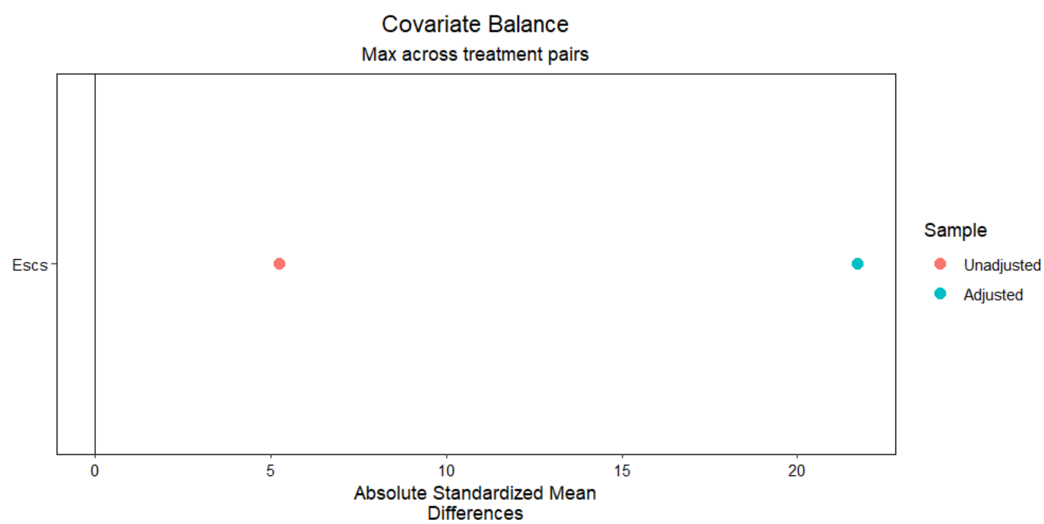

Figure 3.10: Escs balance before and after adjusting using multinomial logit model, combined procedure

The method based on entropy balancing, which ensures perfect balance on the specified moments of the covariates by minimizing the entropy of the weights, is introduced. In terms of balance, the use of this method has resulted in achieving the balance of the variable Escs as it can be seen in Figure 3.11 and the standardized mean difference is less than 0.05.

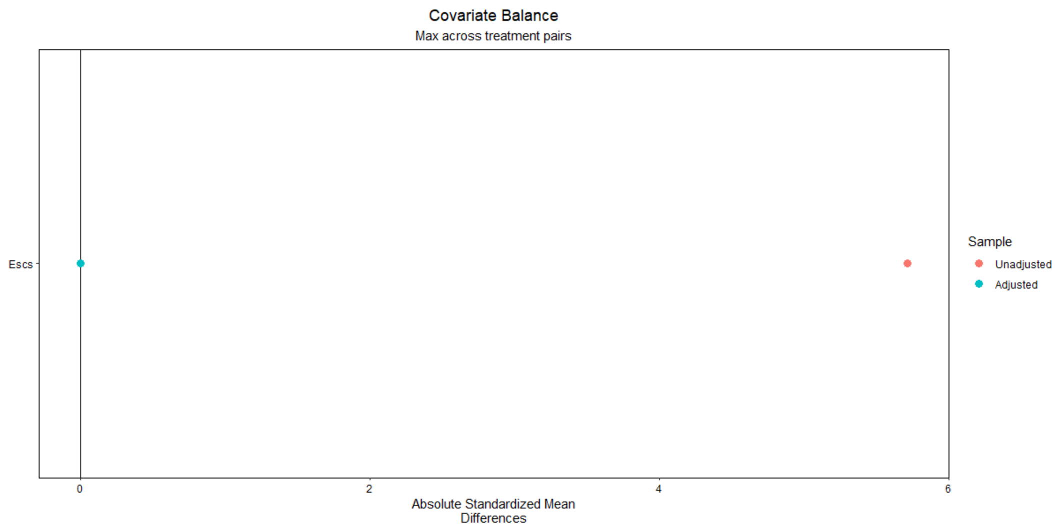

Figure 3.11: Escs balance before and after adjusting using ebal method, combined procedure

Even in the two-step procedure, the best results in terms of balancing are provided by the ebal method. Figure 3.12 shows the results after the two steps.

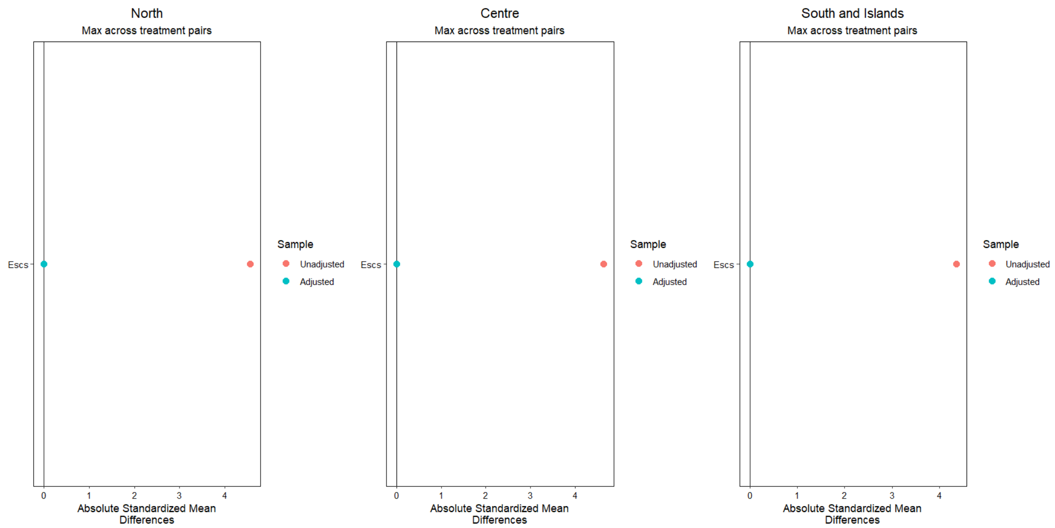

Figure 3.12: Escs balance before and after adjusting using ebal method, two-step procedure

In table 3.7, the ranges of weights assigned for each category of macroregion and high school track with ebal method for the two procedures are provided.

Table 3.7: Weight ranges for each category of macroregion and high school track (ebal method) for the two procedures

	Combined procedure		Two-step procedure	
	Min	Max	Min	Max
North - Humanities	0.013	39.858	0.047	18.111
North - Traditional Science	0.039	23.772	0.108	10.652
North - Applied Science	0.307	17.661	0.483	6.443
North - Other liceo and Technical	0.585	1.564	0.341	3.220
North - Vocational	0.091	27.286	0.045	54.182
Centre - Humanities	0.115	25.881	0.196	14.820
Centre - Traditional Science	0.017	24.787	0.097	8.393
Centre - Applied Science	0.062	18.842	0.403	2.973
Centre - Other liceo and Technical	0.921	1.073	0.156	6.869
Centre - Vocational	0.046	14.905	0.014	28.950
South and Islands - Humanities	0.145	15.724	0.101	22.308
South and Islands - Traditional Science	0.147	10.580	0.074	20.258
South and Islands - Applied Science	0.466	7.590	0.323	18.046
South and Islands - Other liceo and Technical	0.149	5.404	0.552	1.742
South and Islands - Vocational	0.014	20.991	0.036	11.316

Finally, the effective sample size (ESS) was shown in Table 3.8 for each treatment after the balancing based on ebal method for both procedures. This measure gives a sense of how much information remains in the weighted sample.

Table 3.8: Effective sample size with ebal method for the two procedures

	Combined procedure		Two-step procedure	
	Unweighted	Weighted	Unweighted	Weighted
North - Humanities	115.259	40.995	111.030	93.010
North - Traditional Science	310.780	236.450	291.770	494.050
North - Applied Science	276.325	354.985	266.640	569.620
North - Other liceo and Technical	1021.419	1867.576	986.620	1730.290
North - Vocational	373.069	283.559	386.300	121.180
Centre - Humanities	100.966	44.943	95.080	99.400
Centre - Traditional Science	179.512	110.608	168.550	300.920
Centre - Applied Science	121.885	91.569	112.040	234.840
Centre - Other liceo and Technical	529.032	975.716	512.850	716.660
Centre - Vocational	162.048	114.981	164.320	49.100
South and Islands - Humanities	192.102	223.587	196.700	154.680
South and Islands - Traditional Science	355.769	598.589	362.750	367.500
South and Islands - Applied Science	269.240	441.764	272.500	299.000
South and Islands - Other liceo and Technical	1067.929	1682.237	1069.010	1990.650
South and Islands - Vocational	319.903	249.965	317.540	440.910

The combined procedure has a slightly lower total sum of differences than the two-step procedure. This could indicate that the weighted estimates of the combined procedure are closer to the unweighted estimates than the two-step procedure. Furthermore, the combined procedure produces less variability in the weights than the two-step procedure. In conclusion, although there is more variability in the weights, the advantages in the weighted estimates lead to the two-step procedure being chosen and preferred.

3.6 The logit model

After the balancing procedure, a logit model is used to study the effect of treatment (high school track and macroregion) on university enrolment. So, the differences between the various treatment groups are examined to determine the effects of factors considered on university enrolment.

Both procedures try to provide a good balance of the characteristics of the groups. Now, the goal is to investigate the factors that influence the high school-university transition. To reach this goal, we estimate three logit regression models where the response variable is the enrolment at university and the predictors are the socio-economic status, the high school track and the macroregion:

- in the first "*Meta-analytical model*", we use the database obtained from the balancing meta-analytical method in which we eliminated the observations (high school tracks) that caused this unbalance
- in the second "*MPS model*", we use the "weighted" database based on the bal-

ancing method based on multilevel propensity score where did not eliminated any observation

- in the third "*Unbalanced model*", we use the original unbalanced data.

The results of the three models are as follows (Table 3.9):

Table 3.9: Parameter estimates of the models

Covariates		Meta-analytical model Estimate	St. Err.	p-value	MPS model Estimate	St. Err.	p-value	Unbalanced model Estimate	St. Err.	p-value
Intercept		-0.957	0.148	< 0.001	-0.888	0.023	< 0.001	-1.150	0.117	< 0.001
Macroregion	Centre	0.522	0.244	0.032	0.125	0.031	< 0.001	-0.040	0.167	0.809
(North)	South and Islands	0.986	0.187	< 0.001	0.980	0.030	< 0.001	0.283	0.141	< 0.001
High school track	Traditional Science	0.871	0.181	< 0.001	0.844	0.024	< 0.001	0.227	0.130	0.081
(Humanities)	Applied Science	1.220	0.165	0.000	1.552	0.031	< 0.001	0.541	0.139	< 0.001
	Other liceo and Technical	0.063	0.148	0.669	0.240	0.024	< 0.001	0.319	0.125	0.010
	Vocational	0.846	0.161	< 0.001	0.370	0.031	< 0.001	1.640	0.145	< 0.001
Escs		5.329	0.048	< 0.001	3.077	0.015	< 0.001	1.669	0.054	< 0.001
Macroregion * High school track	Centre * Traditional Science	-0.876	0.308	0.004	-0.104	0.038	0.006	0.066	0.203	0.746
(North * Humanities)	South and Islands * Traditional Science	-0.634	0.224	0.005	-0.392	0.036	< 0.001	0.100	0.171	0.560
	Centre * Applied Science	-0.567	0.276	0.040	-0.368	0.052	< 0.001	0.061	0.228	0.788
	South and Islands * Applied Science	-0.646	0.215	0.003	-0.879	0.046	< 0.001	0.041	0.188	0.828
	Centre * Other liceo and Technical	-0.416	0.245	0.090	-0.066	0.034	0.051	0.181	0.188	0.336
	South and Islands * Other liceo and Technical	-0.718	0.189	< 0.001	-0.788	0.032	< 0.001	0.143	0.157	0.362
	Centre * Vocational	-0.433	0.268	0.106	-0.091	0.039	0.020	0.084	0.213	0.692
	South and Islands * Vocational	-0.891	0.214	< 0.001	-0.889	0.037	< 0.001	-0.087	0.175	0.618

Baseline categories are in brackets
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The effect of socio-economic level (Escs) is statistically significant in all three models, indicating that a higher socio-economic level is associated with a higher probability of being enrolled in university. It should be noted that in the unbalanced model, the effect of Escs, although significant, is lower than the estimates in the balanced models. The interactions (first-order effects) between the macroregion and the high school track show how the effect of the macroregion can vary according to the high school track. In the meta-analytical method and MPS models, significant first-order effects between the macroregion and the high school track provide crucial information about the joint effect of these two factors on university enrolment, as they reflect the complexity of students' decision-making process regarding university enrolment. For example, the significant first-order effect between 'Centre' (macroregion) and 'Traditional science' (high school track) means that the effect of macroregion on the probability of enrolling in university varies depending on whether a student attended a traditional science high school in the centre macroregion or in other macroregions. In contrast, there are no statistically significant first-order effects in the unbalanced model. In the meta-analytical and MPS models, balancing the Escs have helped to better capture the complex interactions between regional context, educational pathway and university enrolment.

3.7 Conclusions

Education affects people and society for the rest of their lives. In this context, the transition from school to university is crucial. This phenomenon is influenced by many factors. However, very little research has been carried out in Italy on how these characteristics affect academic outcomes, apart from surveys or small-scale analyses focused on a specific school. The aim of the chapter is to analyse and identify the factors that influence the transition from high school to university in Italy in 2019-2020. There are two major issues to achieving this goal.

The first problem relates to data availability. This was addressed by looking at three data sources and combining them. The three data sets determined the aggregated nature of the combined database. We included 6,763 high school tracks. Among the available variables, macroregion, high school track and socio-economic status are studied as factors of interest in the study of the school-university transition. Combining data from different studies can lead to some biases, among which the one addressed by the methodology presented is the selection bias (second problem). This chapter compares two methods of balancing the quantitative variable of socio-economic status of students. The socio-economic status is balanced with respect to the groups created, taking into account the modalities of the macroregion and the variables of the high school track.

We compare two methods: a meta-analytical method based on the non-parametric comparison of ECDFs and a multilevel propensity score approach. The meta-analytical method is based on the non-parametric comparison of ECDFs between subgroups. The meta-analytical method is divided into two steps: the global combinability holds when the randomization process holds with respect to the levels of macroregion and, then the local combinability holds if the randomisation process holds with respect to the levels of high school track given a category of macroregion. The studies that lead to unbalance are then removed using an algorithm based on two subset selection criteria. The first exists in the literature, the second is proposed here. The final choice is a mixture of the two subset selection criteria to guarantee a trade-off between combinability and number of observations. The meta-analytical method does not depend on the correct specification of a model because it is based on the non-parametric comparison of ECDFs. On the other hand, eliminating observations results in an artificially balanced population rather than the original population. The second method used is based on the concept of multilevel propensity scores. The method used assigns a weight to each observation that increases the similarity between the groups. Again, we propose two procedures for this method by following the same reason line of meta-analytical method. Multilevel propensity score weighting has several advantages and disadvantages. The advantages relate to balance, management of hierarchical structures and flexibility. As with and other weighting methods, the weights applied to subjects do not represent a real population, but an artificially balanced population. The presence of extreme weights can increase the variance of the estimates. In addition, the method depends on the correct specification of the model.

The analysis of the final binomial-logit models (meta-analytical, MPS and unbalanced) provides important insights into the effects of socio-economic status, macroregion and high school track on the transition from high school to university in Italy. First, socio-economic status emerges as a crucial determinant of university enrolment, confirming existing theories in educational sociology and economics (Checchi and Flabbi, 2007; Ballarino and Panichella, 2014; Giancola and Salmieri, 2020; Panichella and Triventi, 2014). The models consistently show that higher socio-economic status is associated with a greater likelihood of enrolling in university, with substantial and statistically significant effects in all three models. This confirms the well-established understanding that socio-economic advantages play a central role in shaping educational trajectories, especially in the balanced models, with students from more affluent backgrounds more likely to pursue higher education opportunities. The joint effects of macroregion and high school track on the transition from high school to university in Italy only become apparent and interpretable through the application of the proposed adjustment methods, which effectively mitigate potential selection biases. These findings highlight the complex interaction between regional context and educational pathways in shaping university enrolment patterns (Checchi and Flabbi, 2007; Ballarino and Panichella, 2014; Panichella and Triventi, 2014).

Both methods seem to achieve good results in terms of balance, but there are some differences in terms of the number of high school tracks/students (studies/patients). In fact, the first meta-analytical method based on the comparison of ECDFs removes observations (high school tracks) to achieve balance, while the second method, based on multilevel propensity scores, does not remove any observations. A possible development of the multilevel propensity score method could be to model a mixed-effects multinomial model (the response is the school track, socio-economic status acts as an explanatory variable and macroregion acts as a random effect) rather than a fixed-effects multinomial model, as is done in both procedures (combined and two-step). The direction of future study is to carry out simulations of the balancing methods. Such investigations would be useful to assess how well these techniques work in different contexts and to provide insightful advice on how best to use them in observational research.

Chapter 4

Conclusions, limitations, and future work

4.1 Conclusions

Higher education, and in particular university education, is a complex system influenced by institutional frameworks, historical legacies and external factors. Individuals pass through many steps called "transitions" during their entire educational journey, including university education.

In Chapter 1, we provide a literature review on transition in higher education and statistical methods for analysing these transitions (longitudinal data analysis). These methods track changes and developments over time, providing valuable insights into the factors that influence educational outcomes and trajectories, and allowing for a better understanding of the interplay between individual, contextual, and temporal elements. There exist challenges associated with the study of transitions, including data quality and availability. Given these challenges and the complexity of transitions in higher education, the thesis focuses specifically on two key pathways: the academic career transitions of faculty members and the high school-university transition of students. One of the main aims of this thesis is to investigate how individual variability in responses to educational interventions, socio-economic factors and environmental conditions during these two pathways influence transitions.

In Chapter 2, we deal with academic transitions for faculty members within Italian universities over the last 20 years, using data from the archives of the Ministry of Universities and Research. Using an event-history analysis approach, it shows the temporal differences in terms of gender, macroregion and field of study in the two transitions (from assistant to associate and from associate to full professor). For both transitions, women take about one and a half years longer than men to move from assistant to associate professor and from associate to full professor. The gender gap is more pronounced in STEM fields and in the southern regions of Italy. The results, especially the gender gap, are also confirmed by the sociological literature.

The analysis of the second pathway is provided in the Chapter 3. We deal with the high school-university transition in Italy, in order to identify the factors that influence it. As anticipated in Chapter 1, the chapter highlights issues related to data availability that we address with a data integration approach. The problem of unbalance in the gerarchical data is addressed using two methods, and some methodological improvements are compared to adjust for socio-economic status across groups: a meta-analytic approach based on non-parametric comparison of ECDFs and a multilevel propensity score method. The first methodological improvement is the subset selection criterion in the meta-analytic method, while the second one is the two-stage procedure in the multilevel propensity score method. Results from binomial logit models highlight the significant impact of socio-economic status, macroregion, and high school track on university enrolment, especially in the balanced models. Individuals with higher socio-economic status are more likely to enroll, while the interaction between macroregion and high school track underlines the complexity of regional and educational pathways. Both methods seem to achieve good results in terms of balance with some differences in terms of dimension (number of observations) and results. Although both methods are based on a fictitious population, the meta-analytic method seems to be preferable, being more flexible in terms of initial assumptions.

4.2 Limitations and Future work

The limitations of the work must be emphasised. There is a presence of partial information. In particular, the data are incomplete with regard to faculty career transitions, as they do not take into account important factors such as age, academic productivity or periods of maternity leave (protected by privacy). This data gap prevents a full understanding of the dynamics of faculty careers. On the other hand, it is possible to work with micro-data for the first time. These microdata are only available for those who enrol at university. For those who do not enrol, we have taken into account the aggregate data from high schools. There is a lack of data on Italian students enrolling in or transferring to foreign universities. This omission leads to an underestimation of the overall transition rate, as it does not capture the full extent of educational mobility.

Future developments could include research and analysis of other educational transitions. Moreover, future studies should focus on developing advanced statistical methods to improve the analysis of transitions in education. For instance, to analyse academic transitions we could consider a multi-state model or a competing risks model (to take into account multiple transitions across careers together). This approach would allow for different analysis by capturing the interdependencies between different types of transitions. The multilevel propensity score method used to analyse the transition from high school to university could be improved by using the multinomial mixed effects model to better account for the hierarchical structure of educational data (e.g. students nested within schools). Moreover, according to us, the balancing procedures applied to the data lack an in-depth comparison in terms of pros and cons of the two procedures. A simulation could be a way to get some insights.

Appendix A

Other Tables and Figures

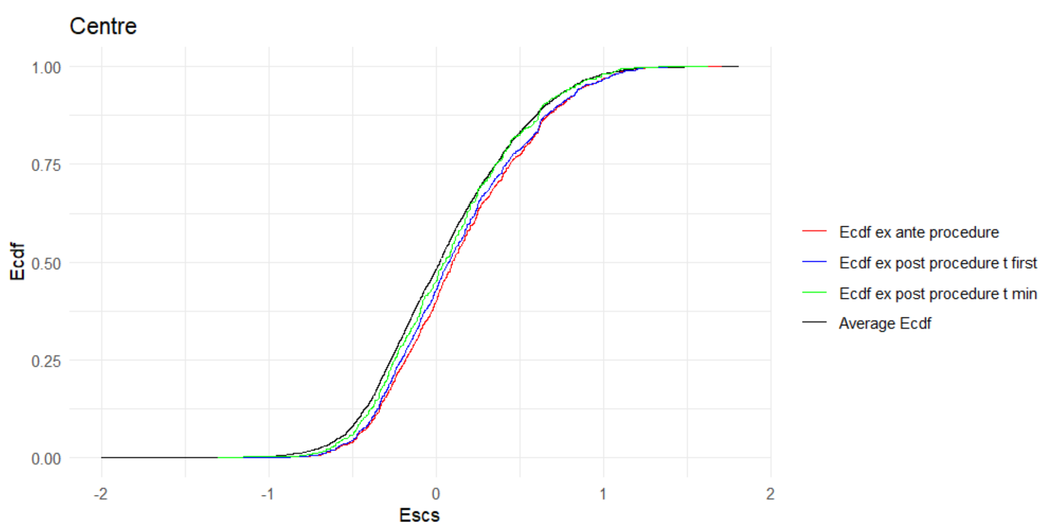

*Figure A.1: ECDFs of the **global** ex ante and ex post combinability procedures for the Centre macroregion*

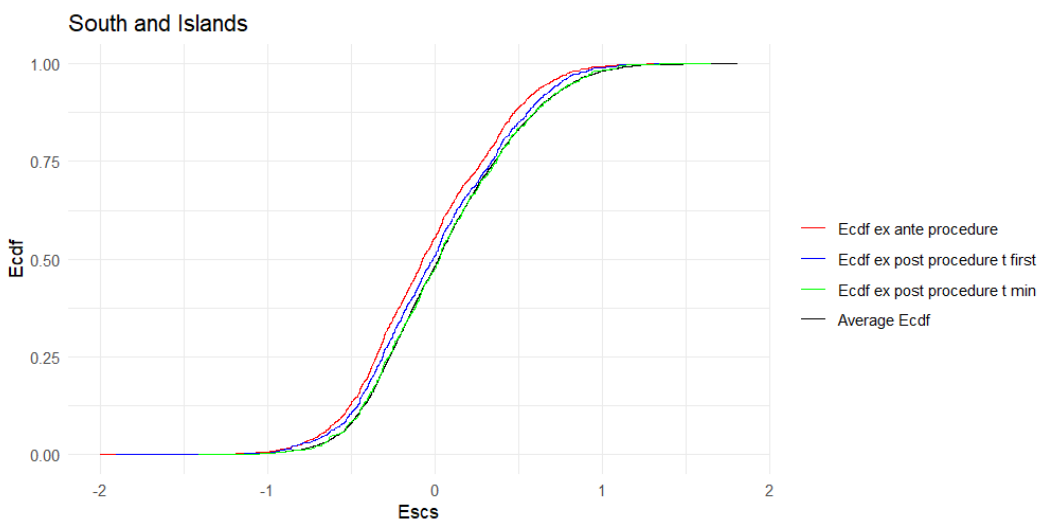

*Figure A.2: ECDFs of the **global** ex ante and ex post combinability procedures for the South and Islands macroregion*

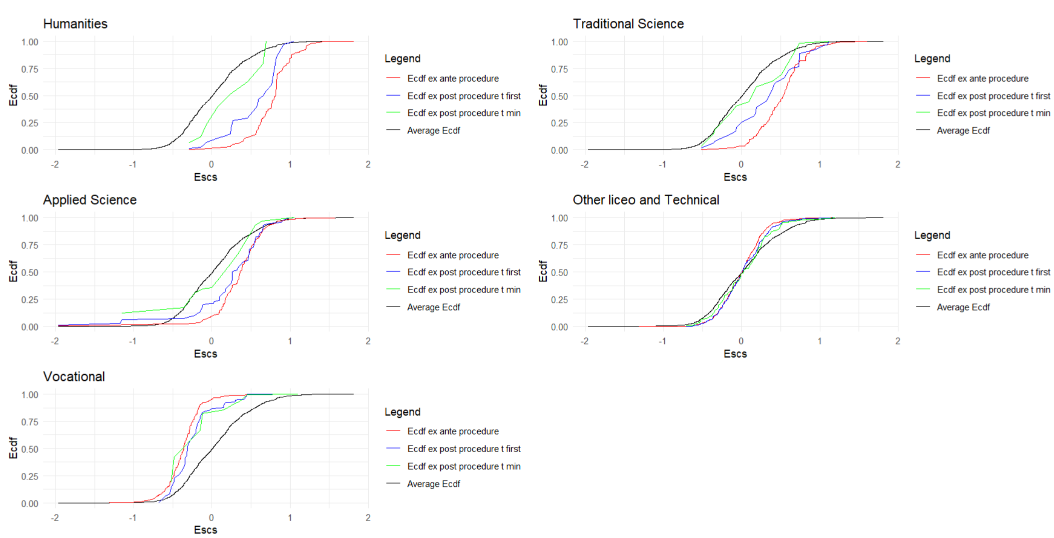

*Figure A.3: ECDFs of the **local** ex post combinability procedures for the North macroregion*

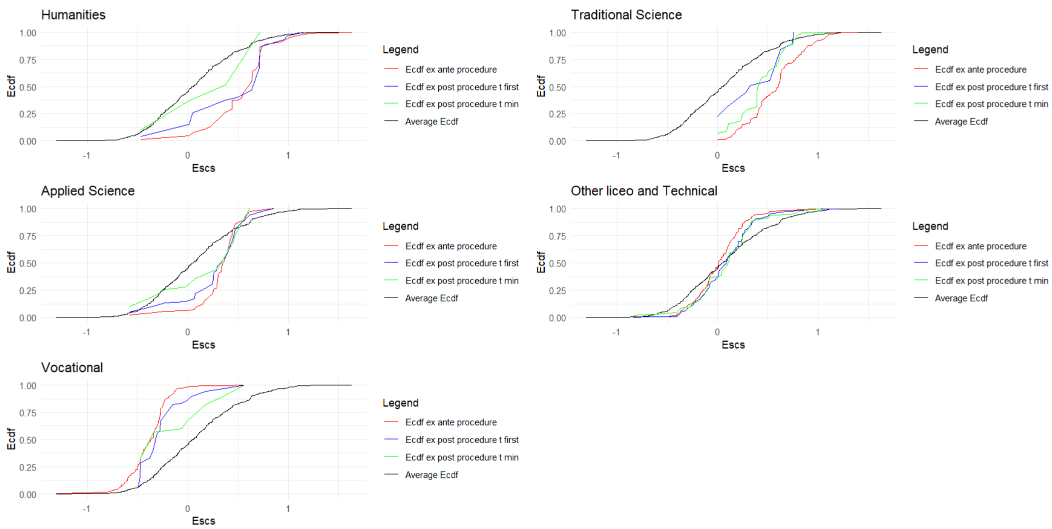

Figure A.4: ECDFs of the ***local*** *ex post combinability procedures for the Centre macroregion*

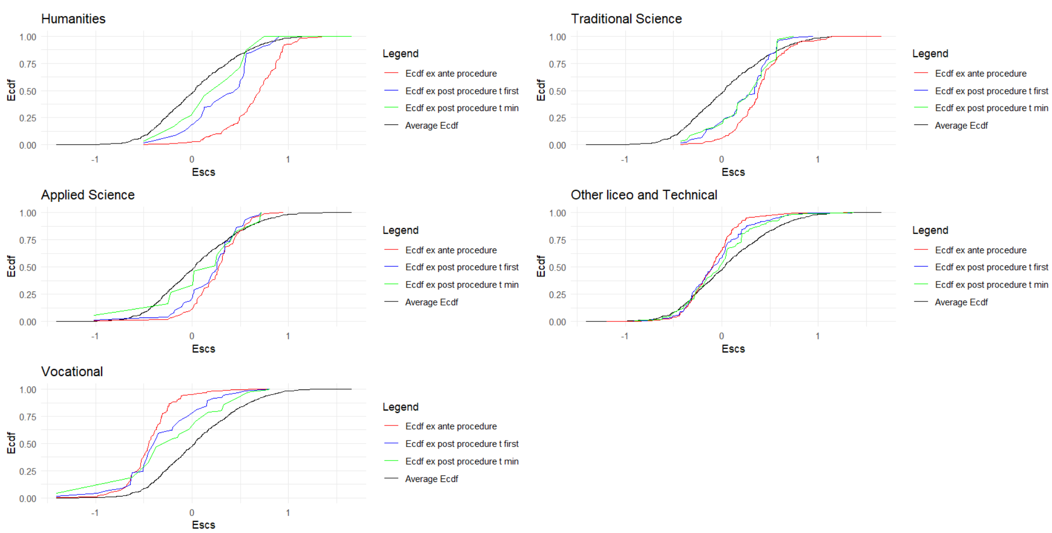

Figure A.5: ECDFs of the ***local*** *ex post combinability procedures for the South and Islands macroregion*

Appendix B

Publications

Scientific publications associated with the thesis research are listed in the following, together with CRediT Author Statement and other publications.

1. Cuntrera, D., Falco, V., Attanasio, M. (2021). Women's career discrimination in Italian Academia in the last 20 years. In: Perna C., Salvati N. & Schirripa Spagnolo F, Book of short papers – SIS 2021, pp. 1493-1498, Pearson, ISBN: 9788891927361

2. Falco, V., Genova, V. G., Priulla, A., Attanasio, M. (2021) L'aumento degli immatricolati nel Mezzogiorno d'Italia nell'anno accademico 2020-21: merito del Covid? *Neodemos*, *https://www.neodemos.info/2021/09/24/laumento-degli-immatricolati-nel-mezzogiorno-ditalia-nellanno-accademico-2020-21-merito-del-covid/*

3. Cuntrera, D., Falco, V., Giambalvo, O. (2022). On the Sampling Size for Inverse Sampling. *Stats*, 5(4), 1130-114. DOI: 10.3390/stats5040067

4. Falco, V., Cuntrera, D., Attanasio, M. (2023) Gender differences in career advancements in Italian universities over the last 20 years. *Genus*, 79, 14 (2023). DOI: 10.1186/s41118-023-00189-7

CRediT Author Statement

In accordance with the Contributor Roles Taxonomy (CRediT), the following section outlines the specific contributions of each author to this thesis:

- **Chapter 1**:
  - Vincenzo Falco: conceptualization, investigation, and writing - original draft.

  - Massimo Attanasio: conceptualization, supervision, writing - review, project administration, and funding acquisition.

- **Chapter 2**:
  - Vincenzo Falco: conceptualization, methodology, software, formal analysis, investigation, data curation, writing - original draft, and visualization.
  - Daniele Cuntrera: conceptualization, methodology, software, formal analysis, investigation, data curation, writing - original draft, and visualization.
  - Massimo Attanasio: conceptualization, supervision, writing - review, project administration, and funding acquisition.

- **Chapter 3**:
  - Vincenzo Falco: conceptualization, methodology, software, formal analysis, investigation, data curation, writing - original draft, and visualization.
  - Martina Vittorietti: conceptualization, methodology, software, formal analysis, investigation, supervision, and writing - review.
  - Massimo Attanasio: conceptualization, 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.

Other publications

- Scaturro, D., Antimo, M., Rizzo, S., Sanfilippo, V., Giustino, V., Messina, G., Martines, F., Falco, V., Cuntrera, D., Iolascon, G., Mauro, G. L. (2021). Effectiveness of rehabilitative intervention on pain, postural balance, and quality of life in women with multiple vertebral fragility fractures: a prospective cohort study. *Journal of Functional Morphology and Kinesiology*, 6 (1). DOI: 10.3390/jfmk6010024

- Scaturro, D., De Sire, A., Terrana, P., Curci, C., Vitagliani, F., Falco, V., Cuntrera, D., Iolascon, G., Mauro, G. L. (2021). Early Denosumab for the prevention of osteoporotic fractures in breast cancer women undergoing aromatase inhibitors: A case-control retrospective study.
textitJournal of Back and Musculoskeletal Rehabilitation, vol. Pre-press, no. Pre-press, 1-6, DOI: 10.3233/BMR-210012

- Scaturro, D., Vitagliani, F., Terrana, P., Cuntrera, D., Falco, V., Tomasello, S., Letizia Mauro, G. (2021). Intra-Articular Hybrid Hyaluronic Acid Injection Treatment in Overweight Patients with Knee Osteoarthritis: A Single-Center, Open-Label, Prospective Study. *Applied Sciences*, 11, 8711. DOI: 10.3390/app11188711

- Falco, V., Genova, V. G., Priulla, A., Attanasio, M. (2021) L'aumento degli immatricolati nel Mezzogiorno d'Italia nell'anno accademico 2020-21: merito del Covid? *Neodemos*, https://www.neodemos.info/2021/09/24/laumento-degli-immatricolati-nel-mezzogiorno-ditalia-nellanno-accademico-2020-21-merito-del-covid/

- Scaturro, D., Costantino, C., Terrana, P., Vitagliani, F., Falco, V., Cuntrera, D., Sannasardo, C. E., Vitale, F., Letizia Mauro, G. (2021). Risk Factors, Lifestyle and Prevention among Adolescents with Idiopathic Juvenile Scoliosis: A Cross Sectional Study in Eleven First-Grade Secondary Schools of Palermo Province, Italy. *International Journal of Environmental Research and Public Health*, 18 (23), URL: 1660-4601/18/23/12335, ISSN: 1660-4601, DOI: 10.3390/ijerph182312335

- Mancuso, S., Scaturro, D., Santoro, M., Di Gaetano, G., Vitagliani, F., Falco, V., Siragusa, S., Gonnelli, S., Letizia Mauro, G. (2021). Bone damage after chemotherapy for lymphoma: a real-world experience. BMC Musculoskelet Disorders, 22, 1024, DOI: 10.1186/s12891-021-04904-3

- Scaturro, D., Vitagliani, F., Terrana, P., Tomasello, S., Falco, V., Cuntrera, D., Spoto, I., Midiri, M., Letizia Mauro, G. (2022). Hybrid Hyaluronic Acid versus High Molecular Weight Hyaluronic Acid for the Treatment of Hip Osteoarthritis in Overweight/Obese Patients. *Journal of Functional Morphology and Kinesiology*, 7 (1), 20. DOI: 10.3390/jfmk7010020

- Scaturro, D., Vitagliani, F., Falco, V., Tomasello, S., Lauricella, L., Letizia Mauro, G. (2022). The Role of Acetyl-Carnitine and Rehabilitation in the Management of Patients with Post-COVID Syndrome: Case-Control Study. *Applied Sciences*, 12 (8)

- Albano, G. D., Rifiorito, A., Malta, G., Sorrentino, E. S., Falco, V., Firenze, A., Argo, A., Zerbo, S. (2022). The Impact on Healthcare Workers of Italian Law n. 24/2017 "Gelli–Bianco" on Patient Safety and Medical Liability: A National Survey. *International Journal of Environmental Research and Public Health*, 19 (14), DOI: 10.3390/ijerph19148448

- Pennisi, G., Enea, M., Falco, V., Aithal, G.P., Palaniyappan, N., Yilmaz, Y., Boursier, J., Cassinotto, C., De Lédinghen, V., Chan, W. K., Mahadeva, S.,

Eddowes, P., Newsome, P., Karlas, T., Wiegand, J., Wong, V.W., Schattenberg, J. M., Labenz, C., Kim, W., Lee, M. S., Lupsor-Platon, M., Cobbold, J. F. L., Fan, J. G., Shen, F., Staufer,, K., Trauner, M., Stauber, R., Nakajima, A., Yoneda, M., Bugianesi, E., Younes, R., Gaia, S., Zheng, M. H., Cammà, C., Anstee, Q. M., Mózes, F. E., Pavlides, M., Petta, S. (2023). Noninvasive assessment of liver disease severity in patients with nonalcoholic fatty liver disease (NAFLD) and type 2 diabetes. *Hepatology*, 78 (1), 195 – 211. DOI: 10.1097/HEP.0000000000000351

- Zerbo, S., La Mantia, M., Malta, G., Albano, G. D., Rifiorito, A., Manco, V., Falco, V., Argo, A. (2023). Mortality of hospitalized patients with sars-cov-2 infection in university tertiary care of italy. *Euromediterranean Biomedical Journal*, 1 (2), 84-90. DOI: 10.3269/1970-5492.2023.18.18

- Filippi, G., Falco, V., (2024). I risultati delle prove TOLC 2023. CISIA (www.cisiaonline.it), ISBN: 9788894356731

- Angelo, T., Di Buono, G., Buscemi, S., Mantica, G., Falco, V., Palacios, V. H., Verri, P., Antelo, R. A., Rosas-Nava, J. E., Crisan, N., Andras, I., Medas, F., Amato, G., Romano, G., Breda, A., Agrusa, A., i-CLARAS (intraoperative Complication in Laparoscopic Renal and Adrenal Surgery) Research Collaborative Study Group (2024). Evaluation of predictive factors for i-CLARAS (intraoperative complications in laparoscopic renal and adrenal surgery): a multicentre international retrospective cohort study. *Scientific reports*, 14 (1). DOI: 10.1038/s41598-024-51696-2

Bibliography

Aalen, O., Borgan, , and Gjessing, H. (2008). *Statistical Analysis of Multi-State Models: The R Package mstate*. Springer.

Abramo, G., Aksnes, D. W., and D'Angelo, C. A. (2021). Gender differences in research performance within and between countries: Italy vs norway. *Journal of Informetrics*, 15(2):101–144.

Acker, S. and Feuerverger, G. (1996). Doing good and feeling bad: The work of women university teachers. *Cambridge Journal of Education*, 26:401–422.

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

Aiello, F., Attanasio, M., and Tinè, F. (2011). Assessing covariate imbalance in meta-analysis studies. *Statistics in Medicine*, 30:2671—2682.

Allison, P. D. (1982). Discrete-time methods for the analysis of event histories. *Sociological Methodology*, pages 61–98.

Allison, P. D. (1984). Event history analysis: Regression for longitudinal event data. *Sage Publications*.

Alper, J. (1993). The pipeline is leaking women all the way along. *Science*, 206(5106):409–411.

Alstadæ ter, A. (2009). Measuring the consumption value of higher education. CESifo (Center for Economic Studies and ifo Institute) Working Paper Series 2799, CESifo.

Andersen, P. K. and Keiding, N. (2002). Multi-state models for event history analysis. *Statistical methods in medical research*, 11(2):91–115.

Andersen, P. K. and Keiding, N. (2012). Multistate models and competing risks analysis. *Statistical Methods in Medical Research*, 21(6):661–665.

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*, 61:309–323.

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

Attanasio, M. and Priulla, A. (2020). *Chi rimane e chi se ne va? Un'analisi statistica della mobilità universitaria dal Mezzogiorno d'Italia*. Franco Angeli Editore.

Austin, P. C. (2011). 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. and Stuart, E. A. (2015). Moving towards best practice when using inverse probability of treatment weighting (iptw) using the propensity score to estimate causal treatment effects in observational studies. *Statistics in Medicine*, 34(28):3661—-3679.

Baldridge, J. V., Curtis, D. V., Ecker, G., and Riley, G. L. (1978). *Faculty Careers and Work Lives: A Professional Growth Perspective*. Jossey-Bass.

Ballarino, G., Bernardi, F., and Panichella, N. (2016). *Education systems and inequality based on social origins: The impact of school expansion and design*, pages 135–157. Education Systems and Educational Inequalities. Ed. by A. Hadjar and C. Gross, Policy Press, Bristol.

Ballarino, G. and Panichella, G. (2014). Origini familiari, scuola secondaria e accesso all'università dei diplomati italiani, 1995-2007. *Scuola Democratica*, 2:365–392.

Ballarino, G. and Panichella, G. (2016). Social stratification, secondary school tracking and university enrollment in italy. *Contemporary Social Science*.

Barnett, R. (2015). *Understanding the University: Institution, Idea, Possibilities*. Routledge.

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., 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.

Barrett, L. and Barrett, P. (2011). Women and academic workloads: Career slow lane or cul-de-sac? *Higher Education*, 61(2):141–155.

Barrett, R., Kuzawa, C. W., and McDade, T. (1998). Epidemiological transitions: complexities and historical context. *The Lancet*, 351(9105):578–579.

Bataille, P., Le Feuvre, N., and Kradolfer Morales, S. (2017). Should i stay or should i go? the effects of precariousness on the gendered career aspirations of postdocs in switzerland. *European Educational Research Journal*, 16(2-3):313–331.

Bauer, K. and Bennett, L. (2003). Applying the principles of significant learning in the e-learning environment. *New Directions for Adult and Continuing Education*, 2003(100):15–23.

Baxter, J. and Wright, E. O. (2000). The glass ceiling hypothesis: A comparative study of the united states, sweden, and australia. *Gender Society*, 14:275–294.

Bean, J. P. and Metzner, B. S. (1985). A conceptual model of nontraditional undergraduate student attrition. *Review of Educational Research*, 55(4):485–540.

Becker, G. S. (1975). *Human Capital*. Chicago University Press, Chicago.

Becker, W. E. (1992). *The Economics of American Higher Education*.

Belfield, C., Boneva, T., Rauh, C., and Shaw, J. (2016). Money or fun? why students want to pursue further education. Ifs working papers, Institute for Fiscal Studies.

Bertrand, M., Goldin, C., and Katz, L. (2010). Dynamics of the gender gap for young professionals in the financial and corporate sectors. *American Economic Journal: Applied Economics*, 2(3):228–255.

Bertrand, M. and Hallock, K. (2001). The gender gap in top corporate jobs. *Industrial and Labor Relations Review*, 55:3–21.

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

Blossfeld, H.-P. and Rohwer, G. (2002). *Techniques of event history modeling: New approaches to causal analysis*. Lawrence Erlbaum Associates.

Boden Jr., R. (1999). Gender inequality in wage earnings and female self-employment selection. *Journal of Socio-Economics*, 28(3):351–364.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., and Rothstein, H. R. (2009). *When Does it Make Sense to Perform a Meta-Analysis?*, chapter 40. John Wiley Sons, New York.

Boudon, R. (1974). *Education, Opportunity and Social Inequality*. John Wiley and Sons, New York.

Boudon, R. (1979). *La Logique du social.* Librairie Hachette.

Bourdieu, C. (1966). L'ecole conservatrice. *Revue Francaise de Sociologie*, 7:325–347.

Bourdieu, C. (1977). *Cultural reproduction and social reproduction.* Oxford University Press, New York.

Bourdieu, C. and Passeron, J. C. (1964). *Les Héritiers.* Les Editions de Minuit, Paris.

Bourdieu, C. and Passeron, J. C. (1977). *Reproduction.* Sage.

Bourdieu, P. (2001). Masculine domination. Stanford University Press.

Box-Steffensmeier, J. M., Cunha, R. C., Varbanov, R. A., Hoh, Y. S., Knisley, M. L., and Holmes, M. A. (2015). Survival analysis of faculty retention and promotion in the social sciences by gender. *PloS One*, 10(11).

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.

Breen, R. and Goldthorpe, J. H. (1997). Explaining educational differential. *Rationality and Society*, 9:275–305.

Brown, L. M. and Lee, S. M. (2017). Examining faculty turnover intentions in higher education: A multilevel analysis. *Journal of Higher Education*, 88(5):752–777.

Brown, M. and Williams, K. (2019). Evidence-based policies for student success. *Journal of Higher Education Policy and Management*, 32:45–61.

Brunello, G. and Checchi, D. (2007). Does school tracking affect equality of opportunity? new international evidence. *Economic Policy*, 22:781–861.

Bryk, A. S. and Raudenbush, S. W. (2002). *Hierarchical linear models: Applications and data analysis methods.* Sage.

Chan, K. C., Yam, S. C., and Zhang, Z. (2016). Globally efficient non-parametric inference of average treatment effects by empirical balancing calibration weighting. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(3):673—-700.

Checchi, D. and Flabbi, L. (2005). Intergenerational mobility and schooling decisions in italy and germany. *IZA Discussion Paper 2876.*

Checchi, D. and Flabbi, L. (2007). Intergenerational mobility and schooling decisions in germany and italy: The impact of secondary school tracks. Iza discussion papers, Institute of Labor Economics (IZA).

Chen, Q., Liu, W., and Zhang, H. (2019). Analyzing faculty promotion patterns using longitudinal data: A case study in a research-intensive university. *Higher Education Research & Development*, 38(5):1023–1040.

Cleves, M., Gould, W., Gutierrez, R., and Marchenko, Y. (2004). *An Introduction to Survival Analysis Using Stata*. Stata Press.

Cochran, W. G. and Rubin, D. B. (1973). Controlling bias in observational studies: A review. *Sankhya: The Indian Journal of Statistics, Series A*, 35(4):417—-446.

Collins, L. M. and Lanza, S. T. (2010). *Latent Class and Latent Transition Analysis: With Applications in the Social, Behavioral, and Health Sciences*. Wiley.

Collins, P. H. and Bilge, S. (2020). *Intersectionality*. John Wiley & Sons.

Collins, R. (1979). *The Credentials Society: An Historical Sociology of Education and Stratification*. Academic Press, New York.

Columbu, S., Porcu, M., and Sulis, I. (2021). 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.

Commission, E. (2021). Eu 2021 report on gender equality. Technical report.

Commission, E., for Research, D.-G., and Innovation (2021). She figures 2021 : gender in research and innovation : statistics and indicators.

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. (2012). Social-origin inequalities in educational careers in italy. performance or decision effects? Department of economics and statistics cognetti de martiis. working papers, University of Turin.

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

Contini, D. and Zotti, R. (2022). *D. Checchi, T. Jappelli, A. Uricchio (eds.) Teaching, Research and Academic Careers-An analysis of the Interrelations and Impacts*, chapter Do financial conditions play a role in university dropout? New evidence from administrative data, pages 39–70. Springer Cham.

Cotter, D., Hermsen, J., Ovadia, S., and Vanneman, R. (2001). The glass ceiling effect. *Social Forces*, 80(2):655–681.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2):187–220.

David, M. and Woodward, D. (1998). *Negotiating the glass ceiling careers of senior women in the academic world*. Falmer Press, London.

De Angelis, I., Mariani, V., Modena, F., and Montanaro, P. (2016). Immatricolazioni, percorsi accademici e mobilita degli studenti italiani (academic enrolment, careers and student mobility in italy). Occasional paper (354), Bank of Italy.

De Laet, T. and Rubin, D. B. (2008). A unified statistical approach for modeling dropout in longitudinal studies. *Journal of the American Statistical Association*, 103(481):1249–1256.

Di Prete, T. A. and Buchmann, C. (2013). The rise of women: The growing gender gap in education and what it means for american schools. *Russell Sage Foundation*, pages 1–277.

Durodoye, R., Gumpertz, M., Wilson, A., Griffith, E., and Ahmad, S. (2020). Tenure and promotion outcomes at four large land grant universities: Examining the role of gender, race, and academic discipline. *Research in Higher Education*, 61(5):628–651.

Elkinawy, S. and Stater, M. (2011). Gender differences in executive compensation: variation with board gender compensation and time. *Journal of Economics and Business*, 63:23–45.

Eurostat (2020). Tertiary education statistics.

Feinstein, A. R. (1995). Meta-analysis: Statistical alchemy for the 21st century. *Journal of Clinical Epidemiology*, 48:71—-79.

Filandri, M. and Pasqua, S. (2019). Gender discrimination in academic careers in italy. page Working papers 201921.

Fine, J. P. and Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, 94(446):496–509.

Forum, W. E. (2021). Global gender gap report 2021. Technical report.

Fox, M. F. (2005). Gender, family characteristics, and publication productivity among scientists. *Social Studies of Science*, 35(1):131–150.

Gaiaschi, C. and Musumeci, R. (2020). Just a matter of time? women's career advancement in neo-liberal academia. an analysis of recruitment trends in italian universities. *Social Sciences*, 9(9).

Gambetta, D. (1987). *Were They Pushed or Did They Jump?: Individual Decision Mechanisms in Education*. Cambridge University Press.

Giambona, F., Porcu, M., and Sulis, I. (2017). Students mobility: Assessing the determinants of attractiveness across competing territorial areas. *Social Indicators Research: An International and Interdisciplinary Journal for Quality-of-Life Measurement*, 133(3):1105–1132.

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

Gibelman, M. (2003). So how far have we come? pestilent and persistent gender gap in pay. *Social Work*, 48(1):22–32.

Ginther, D. K. and Hayes, K. J. (1999). Gender differences in salary and promotion in the humanities. *American Economic Review*, 89(2):397–402.

Glass, G. V. and Smith, M. L. (1978). *Meta-Analysis of Research on the Relationship of Class-Size and Achievement*. Far West Lab. for Educational Research and Development, San Francisco, California.

Goastellec, G. and Pekari, N. (2013). *U. Teichler et E. A. Höhle (Eds.), The Work Situation of the Academic Profession in Europe: Findings of a Survey in Twelve European Countries*, chapter Gender in the Academia between Differences and Inequalities, Findings in Europe, pages 55–78. Dordrecht: Springer.

Grodsky, E. and Rieglecrumb, C. (2010). Those who choose and those who don't: Social background and college orientation. *The Annals of the American Academy of Political and Social Science*, 627:14–35.

Groeneveld, S., Tijdens, K., and van Kleef, D. (2012). Gender differences in academic careers: Evidence for a dutch university from personnel data 1990-2006. *Equality, Diversity and Inclusion: An International Journal.*

Haidich, A. B. (2010). Meta-analysis in medical research. *Hippokratia*, 14:29—-37.

Hainmueller, J. (2012). Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political Analysis*, 20(1):25—-46.

Heckman, J. J. (2000). Policies to foster human capital. *Research in Economics*, 54:3–56.

Heijstra, T., Bjarnason, T., and Rafnsdóttir, G. L. (2015). Predictors of gender inequalities in the rank of full professor. *Scandinavian Journal of Educational Research*, 59(2):214–230.

Ho, D., Imai, K., King, G., and Stuart, E. A. (2011). Matchit: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software*, 42(8):1—-28.

Hougaard, P. (2000). Analysis of multivariate survival data. *Springer Series in Statistics*.

Jackson, C. H., Sharples, L. D., and Thompson, S. G. (2011). Multistate markov models for disease progression with classification error. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 60(2):269–297.

Jackson, M. (2013). *Determined to Succeed?: Performance versus Choice in Educational Attainment*. Stanford University Press.

Johnson, C. D. and Smith, E. R. (2020). Analyzing educational interventions using growth curve modeling: A practical guide. *Educational Researcher*, 49(1):41–55.

Johnson, L., Smith, A., and Brown, E. (2018). Predicting student dropout using survival analysis: A case study in higher education. *Journal of Educational Data Mining*, 10:25–45.

Jones, A. B., Smith, C., and Johnson, D. (2018). Longitudinal modeling in educational research: Current practices and future directions. *Educational Psychologist*, 53(3):172–191.

Jones, R. and Brown, D. (2020). Faculty career progression: A survival analysis approach. *Educational Research Review*, 15(2):234–248.

Kaminski, D. and Geisler, C. (2012). Survival analysis of faculty retention in science and engineering by gender. *Science*, 335:864—866.

Kirchmeyer, C. (2002). Gender differences in managerial careers: Yesterday, today, and tomorrow. *Journal of Business Ethics*, 37:5–24.

Klein, J. P. and Moeschberger, M. L. (2003). *Survival Analysis: Techniques for Censored and Truncated Data*. Springer Science & Business Media.

Labaree, D. F. (2010). Education, markets, and the public good: The selected works of david f. labaree. *Routledge*.

Lafferty, J., McCallum, A., and Pereira, F. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the Eighteenth International Conference on Machine Learning (ICML)*.

Lancaster, T. (1990). *The Econometric Analysis of Transition Data*. Cambridge University Press.

Li, F., Zaslavsky, A. M., and Landrum, M. B. (2013). Propensity score weighting with multilevel data. *Statistics in medicine*, 32(19):3373—-3387.

Lister, K., Andrews, K., Buxton, J., Douce, C., and Seale, J. (2023). Assessment, life circumstances, curriculum and skills: Barriers and enablers to student mental wellbeing in distance learning. *Frontiers in Psychology*, 14.

Long, J. S., Allison, P. D., and McGinnis, R. (1993). Rank advancement in academic careers: Sex differences and the effects of productivity. *American Sociological Review*, pages 703–722.

Long, J. S. and Fox, M. F. (1995). Scientific careers: universalism and particularism. *Review of Sociology*, 21:45–71.

Lubke, G. H. and Muthén, B. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods*, 10:21–39.

Macarthur, C., Foran, P. J., and Bailar, J. C. r. (1995). Qualitative assessment of studies included in a meta-analysis: Des and the risk of pregnancy loss. *Journal of Clinical Epidemiology*, 48(6):739–747.

Marini, G. and Meschitti, V. (2018). The trench warfare of gender discrimination: evidence from academic promotions to full professor in italy. *Scientometrics*, 115(2):989–1006.

Markoulidakis, A., Taiyari, K., Holmans, P., Pallmann, P., Busse, M., Godley, M. D., and Griffin, B. A. (2023). A tutorial comparing different covariate balancing methods with an application evaluating the causal effects of substance use treatment programs for adolescents. *Health Serv Outcomes Res Method*, 23:115–148.

Marlow, S. and Patton, D. (2005). All credit to men? entrepreneurship, finance, and gender. *Entrepreneurship Theory & Practice*, 29(6):717–735.

Martinez, J., Rodriguez, M., and Garcia, L. (2020). Identifying at-risk students using machine learning: A case study in community college education. *Journal of Learning Analytics*, 7(3):98–115.

Misra, J., Lundquist, J. H., Holmes, E., and Agiomavritis, S. (2011). The ivory ceiling of service work. *Academe*, 97(1):22–26.

Morton, T., Ramirez, N. A., Meece, J., Demetriou, C., and Panter, A. (2018). Perceived barriers, anxieties, and fears in prospective college students from rural high schools. *The High School Journal*, 101(3):155–176.

MUR (2020). *Archivio professori nelle università italiane*. https://cercauniversita.cineca.it/php5/docenti/cerca.php.

Noah, G. (2023). *WeightIt: Weighting for Covariate Balance in Observational Studies*. https://ngreifer.github.io/WeightIt/.

OECD (2018). Equity in education: Breaking down barriers to social mobility. Technical report, Paris.

OECD (2019a). Education at a glance 2019 - oecd indicators 2019. Technical report, Paris.

OECD (2019b). *PISA 2012 Results: Excellence through Equity (Volume II): Giving Every Student the Chance to Succeed*. PISA, OECD Publishing, Paris.

OECD (2020). Education at a glance 2020: Oecd indicators. Technical report, Paris.

Ono, M. and Miller, H. P. (1969). Income nonresponses in the current population survey. *ASA Proc Social Statistics*, pages 277–299.

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*, 58:666—-693.

Parsons, E. and Priola, V. (2013). Agents for change and changed agents: The micropolitics of change and feminism in the academy. *Gender, Work & Organization*, 20(5):580–598.

Pascarella, E. T. and Terenzini, P. T. (2005). *How College Affects Students: A Third Decade of Research*. Jossey-Bass.

Perna, L. W. (2001). Sex and race differences in faculty tenure and promotion. *Research in Higher Education*, 42(5):541–567.

Perna, L. W. (2005). Sex differences in faculty tenure and promotion: The contribution of family ties. *Research in higher Education*, 46(3):277–307.

Perry, R. P. and Smart, J. C. (1997). *The Scholarship of Teaching and Learning in Higher Education: Contributions of Research Universities*. Indiana University School of Education.

Peterson, H. (2016). Is managing academics “women’s work”? exploring the glass cliff in higher education management. *Educational Management Administration & Leadership*, 44(1):122–127.

Pillas, D., Marmot, M., Naicker, K., Goldblatt, P., Morrison, J., and Pikhart, H. (2014). Social inequalities in early childhood health and development: a european-wide systematic review. *Pediatr Res*, 76:418–424.

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.

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.

Raudenbush, S. W. and Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Sage Publications, 2 edition.

Regner, I., Thinus-Blanc, C., Netter, A., Schmader, T., and Huguet, P. (2019). Committees with implicit biases promote fewer women when they do not believe gender bias exists. *Nature Human Behaviour*, 3(11):1171–1179.

Reifeis, S. A. and Hudgens, M. G. (2020). On variance of the treatment effect in the treated using inverse probability weighting.

Rivera, L. A. (2019). Pedigree: How elite students get elite jobs. *Princeton University Press*.

Robins, J. M., Hernán, M. , and Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11(5):550—-560.

Rodriguez-Muñiz, L. J., Bernardo, A. B., Esteban, M., and Díaz, I. (2019). Dropout and transfer paths: What are the risky profiles when analyzing university persistence with machine learning techniques? *PLoS ONE*, 14(6).

Roseberry, L., Remke, R., Klasson, J., and Holgersson, T. (2016). The gender gap in european business schools: A leadership perspective. Technical report.

Rosenbaum, P. R. and Rubin, D. B. (1983). The worldwide expansion of higher education in the twentieth century. *Biometrika*, 70(1):41–55.

Roth, W. D. and Sonnert, G. (2010). The costs and benefits of "red tape": Anti bureaucratic structure and gender inequity in a science organization. *Social Studies of Science*, 41:385–411.

Sax, L. J., Hagedorn, L. S., Arredondo, M., and DiCrisi, F. A. (2002). Faculty research productivity: Exploring the role of gender and family-related factors. *Research in higher education*, 43(4):423–446.

Schofer, E. and Meyer, J. W. (2005). The worldwide expansion of higher education in the twentieth century. *American Sociological Review*, 70(6):898–920.

Scott, W. R. and Meyer, J. W. (1994). *Institutional Environments and Organizations: Structural Complexity and Individualism*. Sage Publications.

Scott-Clayton, J. (2012). *The Shape of the River: Long-term Consequences of Considering Race in College and University Admissions*. Princeton University Press.

Singer, J. D. and Willett, J. B. (2003). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*.

Smith, A., Johnson, B., and Martinez, C. (2018). Dynamics of academic trajectories. *Journal of Higher Education*, 42(3):456–472.

Smith, B. T., Egro, F. M., Murphy, C. P., Stavros, A. G., Kenny, E. M., and Nguyen, V. T. (2019). Change is happening: an evaluation of gender disparities in academic plastic surgery. *Plastic and reconstructive surgery*, 144(4):1001–1009.

Smith, J. K. and Johnson, P. Q. (2019). Factors influencing faculty career transitions in higher education: A review of the literature. *Higher Education Research & Development*, 38(3):528–544.

Stanfield, J. (2011). Impact investment in education. *Economic Affairs*, 31(3).

StataCorp (2019). Stata statistical software: Release 16. College Station, TX: StataCorp LLC.

Steinmetz, S. (2011). The contextual challenges of occupational sex segregation - deciphering cross-national differences in europe. Springer.

Stewart, P., Ornstein, M., and Drakich, J. (2009). Gender and promotion at canadian universities. canadian review of sociology. *Revue Canadienne de Sociologie*, 46(1):59–85.

Stone, P. (2007). *Opting Out?: Why Women Really Quit Careers and Head Home*. University of California Press.

Teixeira, C. T., Silva, C. R., Mesquita, J. S., and Rampazo, A. V. (2018). "assédio sexual no contexto acadêmico da administração": relações de poder cotidianas sobre as quais elas já disseram, e nós vamos dizer mais. E.V CBEO - Curitiba. http://www.sisgeenco.com.br/sistema/cbeo/anais2018/ARQUIVOS-resumos/GT11-197-152-20180520170758.pdf.

Terenzini, P. T. (1999). The study of college impact: Past, present, and future. *Research in Higher Education*, 40(6):605–620.

Thangarajah, J., Kolling, A., and Weibelzahl, S. (2016). Predicting student dropouts in massive open online courses using hidden markov models. In *Proceedings of the European Conference on Technology Enhanced Learning (EC-TEL)*.

Thoemmes, F. J. and Ong, A. D. (2016). A primer on inverse probability of treatment weighting and marginal structural models. *Emerging Adulthood*, 4(1):40—-59.

Thomas, P. A., Diener-West, M., Canto, M. I., Martin, D. R., Post, W. S., and Streiff, M. B. (2004). Results of an academic promotion and career path survey of faculty at the johns hopkins university school of medicine. *Academic Medicine*, 79(3):258–264.

Thun, C. (2020). Excellent and gender equal? academic motherhood and "gender blindness" in norwegian academia. *Gender, Work & Organization*, 27:166—-180.

Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research*, 45:89–125.

Trowman, R., Dumville, J. C., Torgerson, D. J., and Cranny, G. (2007). The impact of trial baseline imbalances should be considered in systematic reviews: a methodological case study. *Journal of Clinical Epidemiology*, 60:1229—-1233.

Tyler, W. (1981). Origins and destinations: Family, class, and education in modern britain. *Journal of Social Policy*, 10(1).

Usala, C., Porcu, M., and Sulis, I. (2023). The high school effect on students' mobility choices. *Statistical Methods & Applications*, 32:1259–1293.

Van de Werfhorst, H. G. and Mijs, J. J. B. (2010). Educational system and social mobility: A comparative study. *European Sociological Review*, 26(3):243–260.

Van den Brink, M. and Benschop, Y. (2012a). Gender practices in the construction of academic excellence: Sheep with five legs. *Organization*, 19(4):507–524.

Van den Brink, M. and Benschop, Y. (2012b). Slaying the seven-headed dragon: The quest for gender change in academia. *Gender, Work & Organization*, 19(1):71–92.

Van Doren, R. C. (1991). *A History of Knowledge: Past, Present, and Future*. Birch Lane Pr.

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.

Vossenberg, S. (2013). Women entrepreneurship promotion in developing countries: What explains the gender gap in entrepreneurship and how to close it? page WorkingPaper No. 2013/08.

Wang, J., Wang, X., and Zhang, D. (2018). Latent transition analysis with covariates: A comparison of multiple-group and separate-group approaches. *Structural Equation Modeling*, 25(2):203–218.

Wang, L. and Brown, C. H. (2009). Bandwidth selection for smoothing spline analysis of variance models. *Journal of the American Statistical Association*, 104(487):1073–1083.

Weisshaar, K. (2017). Publish and perish? an assessment of gender gaps in promotion to tenure in academia. *Social Forces*, 96(2):529–560.

Wirth, L. (2001). *Breaking the throw the glass ceiling. Women in management.* International Labour Office, Geneve.

Wolfinger, N. H. and Wilcox, W. B. (2008). Happily ever after? religion, marital status, gender and relationship quality in urban families. *Social Forces*, 86(3):1311–1337.

Wyarczyc, P. and Renner, C. (2006). The "gender gap" in the scientific labour market: The case of science, engineering and technology-based smes in the uk. *Equal Opportunities International*, 25(8):660–673.

Yap, M. and Konrad, A. (2009). Gender and racial differentials in promotions: Is there a sticky floor, a mid-level bottleneck, or a glass ceiling? *Industrial Relations*, 64.

Zhang, J. and Yuehua, W. (2007). k-sample tests based on the likelihood ratio. *Computational Statistics and Data Analysis*, 51(9):4682–4691.

List of Figures

List of Tables