ORIGINAL PAPER

The high school effect on students' mobility choices

Cristian Usala[1] · Mariano Porcu[1] · Isabella Sulis[1]

Abstract

This paper looks into the relationship between students' university choices and their secondary school background. The main aim is to assess the role of secondary schools in steering university applications toward local or non-local institutions, also in the light of the tertiary education supply available in students' areas of residence. With this aim, we classify students' mobility choices by using a robust definition of local and non-local universities that accounts for the uncertainty in the definition of students' local areas and their characteristics. In this framework, we apply a multilevel model to jointly consider the high school effect on the probability of students belonging to one specific category of mobility (*local*, *forced non-local*, *free non-local*) conditional upon students' macro areas of residence, their chosen university and field of study. The findings highlight that high schools have a relevant role in affecting students' mobility choices, especially when considering local universities. The magnitude of the effect depends on students' macro area of residence. In particular, this result highlights that schools may pursue specific guidance policies to address students' choices toward local universities; furthermore, it suggests that their influence on students is stronger in areas hosting the most important universities.

Keywords Higher education · Mobility choices · Multilevel model · Uncertainty · School

✉ Cristian Usala
  cristian.usala@unica.it

  Mariano Porcu
  mariano.porcu@unica.it

  Isabella Sulis
  isulis@unica.it

[1] Department of Social and Political Sciences, Università degli Studi di Cagliari, Viale S. Ignazio, 78, 09123 Cagliari, Italy

🖄 Springer

1 Introduction

In the last decade in Italy, many researchers have focused their attention on describing university students' mobility flows between different geographical regions by highlighting that mobility choices are strictly associated to several demographic, economic and social features of the Italian contemporary history and society. Indeed, this phenomenon is related to: (1) demographic issues linked to the depopulation of the disadvantaged areas of the country; (2) the so-called brain drain phenomenon, which contributes to depleting the less developed regions; (3) families strategies addressed to ensure better education opportunities and future employment of young generations and to inter-generational mechanisms of reproduction of disparities and inequalities; (4) gender stereotypes and students' aspirations; (5) the consequences of educational policies carried out in the last decade and tertiary education reforms (6) and the effects of the local context in terms of geographical and sociocultural characteristics (Attanasio and Enea 2019; Barone and Assirelli 2020; Ciriaci 2014; D'Agostino et al. 2019b; Impicciatore and Tosi 2019; Santelli et al. 2022). Empirical studies have well documented the presence of an almost unidirectional flow of students from southern to central or northern Italian regions for bachelor's studies. A phenomenon that becomes even harsher in the transition from bachelor's to master degrees (Columbu et al. 2021; Enea 2018) and that may anticipate future migration choices of highly educated individuals (Oggenfuss and Wolter 2019). Moreover, these elements are even more crucial given that students and universities affect regional competitiveness and growth potential (Salter and Martin 2001; Valero and Van Reenen 2019), and may promote intergenerational mobility (Chetty et al. 2020; Chetty and Hendren 2018). Moreover, several studies document that an important share of variability in students' mobility choices is ascribable to divergences in contextual factors and the socioeconomic contexts between the geographical areas of origin and destination (D'Agostino et al. 2019b; Giambona et al. 2017). Santelli et al. (2022) highlight that students attending a *liceo* in the main cities in Campania have a higher propensity to be attracted by the reputation of the universities; moreover, other studies confirm the higher propensity to select non-local universities for students attending schools with the highest profiles in terms of socioeconomic conditions of students' families, namely classical and scientific high schools (Rizzi et al. 2021).

According to the literature, the divergences in students' educational choices can be related to two theoretical paradigms: investment and consumption theories (Attanasio 2022; Foot and Pervin 1983). According to the former, individuals' education choices depend on comparing the expected returns of education and its expected costs, which are functions of both tangible and intangible assets. Meanwhile, according to the consumption theory, the demand for education follows the conventional economic theory, and it depends on the interplay between students' preferences and income, education's relative price and universities' characteristics (Krezel and Krezel 2017). Recent findings provided by Rizzi et al. (2021) suggest that in Italy the mobility of northern first-year

students toward cities such as Milan and Turin is mainly driven by the investment perspective; nevertheless, as highlighted by Attanasio (2022), the Italian framework is characterized by an almost unidirectional flow toward the main northern cities that are also those that can provide students with higher standards in terms of services, infrastructures, leisure activities and opportunities. Indeed, these peculiarities would suggest an intersection of the two perspectives behind students' mobility choices.

Nevertheless, both perspectives see students' decision process as a multi-stage process in which individuals develop their educational aspirations, search information on universities and colleges, and finally choose their university (see for example Hossler and Gallagher 1987). In this decision process, the role of the high school environment is paramount since it affects both the development of students' capabilities as well as the information available to them regarding the universities, which, in turn, determine inter-generational social mobility patterns and the presence of social inequalities.[1] However, despite the proliferation of studies on the phenomenon, the literature is characterized by a relevant gap in understanding the role that high schools play in shaping students' mobility decisions.

Another critical aspect of the previous literature is a high level of heterogeneity in the definition of students' local area and, therefore, students' mobility choices. Indeed, these are usually classified by using *ad hoc* definitions that depend essentially on the aims of the analysis. For example, studies interested in analyzing the pull and push factors that drive students' and families' choices mainly focus on the analysis of the South-North mobility pathways or bound the analysis to flows between macro geographical areas (e.g. Pitzalis and Porcu 2015; Genova et al. 2019). Others focus their attention on the patterns of mobility of students that live in specific areas in order to shed light on educational policies which may counteract the outgoing flows of students by enhancing local universities' attractiveness (e.g. Santelli et al. 2019). In the same frame, studies that investigate university prestige, and its influence on students' university choice, consider student flows as an indirect indicator of university quality and mainly focus on the classification of students as stayers or movers with respect to a given university or a given town (e.g. Giambona et al. 2017).

Our contribution aims to fill these gaps by proposing a robust classification of students' mobility choices and estimating how these choices relate to the high school attended by students.

Concerning students' choices classification, we define our outcome variable based on: (1) the supply of tertiary education in students' local area, (2) the chosen subject of study, and (3) the travel time needed to reach the nearest university. In particular, students' local area is defined as the territory between students' town of residence and the nearest university. This definition allows us to avoid assumptions that rely only on administrative borders or deterministic thresholds, but it rather depends on the availability of universities in each student's residence area. Based on this setting,

---

[1] See for example Engberg and Wolniak (2010) on the effect of the high school context on students' post-secondary enrollment decisions.

we define three choice categories: *local* universities, *forced non-local* universities, and *free non-local* universities. *Local* universities are those placed within students' local areas. *Forced non-local* universities are those placed outside the student's local area but are the nearest university that provides a program in her/his chosen field study. Finally, *free non-local* universities are those hosted in towns farther than the *local* and *forced non-local* universities. This definition allows us to distinguish between two kinds of mobility: one related to students' preferences toward a specific field of study, the other related to students' preferences regarding a specific university/territory in Italy. Indeed, students who choose the *forced non-local* alternative may be more prone to stay in their local area than those who choose the *free non-local* alternative. Therefore, their mobility choices should be considered more similar to those of students who choose the *local* alternatives. Indeed, our results show that students in *local* universities have similar characteristics to those that choose *forced non-local* universities. This difference is significant from a policy point of view since the out migration due to students that choose the *forced non-local* alternatives may be reduced by increasing the supply of degree programs in their local area. Moreover, to account for the uncertainty in the definition of the *local* and *non-local* universities, we generate multiple thresholds for students' local area definition by adding a random amount of travel time to the observed thresholds. Based on this categorization, we apply Multiple Imputation Analysis procedures to assess the sensitivity of the random selection of the thresholds and to combine the results in a single statement.

As for the effect of high schools on students' mobility decisions, we start from the hypothesis that the high school environment, characterized by a set of peers, teachers, and counseling policies, can be treated as a not observable variable that affects students' preferences toward *local* or *non-local* universities. At this aim, we look at the choices of students who have attended the same high school before enrolling at the university and compare students who belong to different schools but have similar socio-demographic characteristics. Nevertheless, there is high variability in the propensity to move across disciplinary fields. These divergences are not just related to the presence of the curricula in the local area but also to the opportunities in terms of lifestyle changes offered by the new destination. Moreover, the propensity to choose a non-local university and the choice of the field of study are also connected to the type of secondary school attended. Thus, to disentangle the school effect in orienting student choices, removing the variability in the propensity to be in mobility connected to these aspects is crucial. To our knowledge, this is the first attempt to use Italian micro-data from the National Student Archive managed by the Italian Ministry of University to shed light on the effect of the secondary school environment on students' university choices in terms of selection of local or non-local universities.

At this aim, a multilevel multinomial approach has been considered to estimate the choice probabilities associated with *local*, *forced non-local* and *free non-local* choices. In particular, we jointly consider the clustering of university applicants in secondary schools and the heterogeneity in the choices between macro geographical areas and the subject of study. This approach allows us to estimate two parameters for each high school that inform on the role that the school plays in affecting students' propensity to choose a non-local university (*free non-local* or *forced*

non-local), holding constant across schools the other sources of heterogeneity related to students' characteristics and geographical peculiarities which can affect students' mobility choices.

The analysis indicates that high schools have a relevant role in affecting students' mobility choices, especially when considering *local* universities. In this case, variance partition coefficient results indicate that, depending on students' macro areas of residence, differences between high schools explain between 24.38% and 53.41% of the variability in students' mobility choices that is not explained by students' characteristics or differences between fields of study. The high school effect is stronger for students that reside in the Islands and central Italy while weaker for those residing in the North of the country. From high school perspectives, the results show that schools with a higher than average propensity to steer students toward *local* universities may have a relevant role in keeping southern students and those residing in the Islands close to their local area and counteract the well known asymmetric flow of students from southern to northern regions of the country.

This paper is structured as follows. Section 2 briefly reviews the relevant literature on students' university choices and peer effects. Section 3 provides information on the higher education system in Italy and describes the data. Section 4 presents the empirical method adopted and explains our analysis strategy. Section 5 presents the main empirical results. Finally, Sect. 6 provides our discussion of the results and conclusions.

2 Setting the background

Among the causes of students' mobility choices, many studies have highlighted how the economic conditions of the territories that host the universities affect students' universities choices, pointing out as the differences in the employment opportunities are one of the main drivers of families' decisions to invest in tertiary education studies in the areas where the return of education can be faster and more profitable (Dotti et al. 2013; Impicciatore and Tosi 2019; Giambona et al. 2017).

Besides economic rewards, many other important factors may influence students' and families' decisions, such as the willingness to travel to reach the university, the distance between students' town of residence and the university, the accessibility of competing institutions, the availability of public means of transports, and policies aimed to financially support students such as scholarships or places in dormitories (Castleman and Long 2016; Cattaneo et al. 2017b; Pigini and Staffolani 2016; Spiess and Wrohlich 2010; Suhonen 2014; Türk 2019). For universities' characteristics, other authors have highlighted the role played by tuition fees (Dwenger et al. 2012; Hübner 2012; Long 2004) and institutions' quality in specific domains related to research, services, teaching, internationalization, and the perception that graduates have of their overall university experience (Biancardi and Bratti 2019; Bratti and Verzillo 2019; Ciriaci 2014).

Cattaneo et al. (2017a) highlight that students' university choices can be rationalized within three main theories: Human Capital Theory, Signalling Theory and Preference Theory. The Human Capital Theory interprets student mobility

from an investment perspective, where individuals' choices concerning investments in education result from a decision process addressed to maximize students' job opportunities. Focusing on the Italian case, where the demand for qualified jobs is lower than the supply, the Signaling Theory assigns a predominant role to the credentials of the universities in terms of which act as a criterion to differentiate the values of students' degrees. Lastly, Preference Theory is much related to the consumption perspective in explaining human migration choices: individuals make their choices based on the lifestyles that the socioeconomic environment of the destination place offers. The prevalence of one of these theories in explaining students' choices depends on a complex interaction between many factors, such as the business cycle and gender-related stereotypes, as documented by Cattaneo et al. (2017a) in analyzing the determinants of students' choices during the economic crisis period.

Recent researches focus on the peculiarities of the internal mobility of students in Italy as a process that contributes in reproducing social inequalities and widens the disparities in the opportunities between southern and northern students or between those who come from less-educated families and those that belong to families with higher educational levels. For example, Impicciatore and Tosi (2019) provide evidence that parental education is one of the main determinants of students' university choices and that the South-North migration flow is strictly linked to high cultural resources of families which perceive the investment in education as the foremost opportunity for reinforcing their social status. In contrast, internal mobility within southern regions is not associated with parental background.

Other studies have recently investigated the hypothesis of the existence of chain migration effects by looking at the role of private information on university experiences shared by communities of students (Genova et al. 2019). The joint use of Social Network Analysis tools and clustering techniques allowed the authors to disentangle the role played by family and friendship ties as the residual effect after accounting for all the push and pull effects related to origin and destination places (Santelli et al. 2022).

Barone and Assirelli (2020) investigated the gender segregation in higher education in Italy by highlighting that peers' behaviours influence student choices regarding the field of study selection. Moreover, individuals' choices are also driven by the school environment, counsellor services, teachers' recommendations and classmates' preferences. These findings suggest that students' university choices are not merely the results of rational choices based on cost-benefit analysis but are mediated by several factors related to student cultural identity and sense of belonging to a community. Focusing on the United States, Engberg and Wolniak (2010) develop a conceptual framework in which the high school's environment is the driver of college choice decisions. Their research shows that the effects of the high school context on university enrollment depend on the school's endowment in terms of human, social and cultural capital resources above and behind the individual characteristics. Using a multilevel approach on longitudinal data, which includes individual and school-level predictors (as compositional variables of individual characteristics), the authors conclude that both peers' and parents' networks have a relevant impact on enrollment decisions in two and four-year programs.

Moreover, a comparison between school-level and student-level coefficients reveals a higher impact on four-year program enrollments of school-level variables, college aspirations of family and friends, parent-to-parent contact, and the number of friends who plan to attend a four-year program.

3 Data

This section describes the main features of the educational system in Italy and the data used in the present analysis. We focus on high school graduates' decisions to continue their studies at the tertiary level in a *local* or *non-local* university, considering the availability of university degree programs in their residence area, their secondary school background, and their characteristics.

3.1 Data sources and eligibility criteria

The analysis relies on the administrative data from the database MOBYSU.IT regarding all the population of Italian high school leavers enrolled in a bachelor's program in an Italian university between 2016 and 2020.[2]

Given our interest in understanding the role of secondary school background in shaping students' mobility choices, we define our data according to a set of eligibility criteria. First, we do not consider the students enrolled in programs with a national entrance test since their location decisions depend on their ranking positions rather than their preferences. Second, we do not consider students enrolled in e-learning universities not requiring students to move to attend their degree program classes. Third, we retain in our data only those freshmen who gained their high school diploma in an Italian institution after 2015. This choice is because we have available information at the school level since the academic year 2016/2017. Therefore, starting from a population of 1,244,934 freshmen enrolled in an Italian university, our population of interest consists of 1,041,755 students. Among these students, we do not consider those that do not report any information on their high school background (18,434 records) or their city of residence (1138 records) and those that have enrolled at university before their high school diploma (818 records). In total, we lose information on 20,390 students. However, they represent only 1.96% of our population of interest. Thus, since our results should not be affected by these students, we advance the hypothesis that they are missing at random.

Following this strategy, our final data consists of 1,021,365 records. For each student, we observe several characteristics regarding the chosen university, their town of residence, and the type of high school they attended that will be described in the following sections.

---

[2] Data drawn from the Italian archive *Anagrafe Nazionale della Formazione Superiore* has been processed according to the research project 'From high school to the job market: analysis of the university careers and the university North-South mobility' funded by the Ministry of University and Research.

3.2 University supply in students' local area

The dataset includes enrollments in 82 universities located in 209 towns in all the Italian regions. Each university provides a set of degree programs that may be classified into three categories: 3-year bachelor's programs, 5- or 6-year degree programs, and masters' degree programs. In this analysis, we focus only on bachelor's programs.[3] Moreover, we consider only universities that for admissions rely on an ex-post screening system in which the only requirement for enrollment is a high school diploma.[4] Following this strategy, we observe 49 different degree programs that can be classified according to the 'International Standard and Classification of Education: Field of education and training' (ISCED-F 2013, see UNESCO Institute for Statistics 2014) in 10 fields of study shown in Table 8 in the Appendix. The ISCED-F 2013 classification is based on the similarities between programs' disciplinary contents. Therefore, it allows us to classify the programs that students may consider as similar or substitutes in the same category. In our analysis, this feature is crucial since we want to distinguish between students that choose a *non-local* university to attend a degree program that would also be available at their *local* university (i.e., closest to their residence), from the ones that are forced to move to attend a course that is not provided in their *local* universities.

We classify students' mobility choices depending on: (1) the supply of universities in their local area, (2) the travel distance between their city of residence and the nearest university, (3) the chosen field of study. Data on travel distances are gathered by considering the minimum travel distance by car between any two pairs of Italian towns available from the National Statistical Institute (ISTAT) and the information from Google Maps.[5] This data has been used to compute the travel distance (in min) between each student's city of residence and all the Italian universities. Table 1 shows, for each Italian region and macro region, the descriptive statistics on the average travel time needed by students residing in the region to reach their chosen university from their city of residence, along with the minimum distance from the nearest university computed considering (a) all the fields of study (column 3) and (b) the field of study chosen by the student (column 4).

As we can note, the average distance students travel depends heavily on the macro region of residence considered. The average distance recorded for students residing in the Islands is 2.5 times the national average, while in the South and North-West it is respectively about 1.5 times and 0.5 times the average distance recorded in the country. This element is related to the well known Italian North–South divide (e.g. Attanasio and Enea 2019), characterized by an almost

---

[3] In the time frame considered, 88.2% of all students enrolled in an Italian university were in a bachelor's degree program.

[4] We do not consider the students enrolled in the 5- or 6-year programs in medicine, veterinary medicine, agronomic, primary education, and architecture. The number of students enrolled in these programs is defined ex-ante and the applicants are evaluated via an entrance test. See Declercq and Verboven (2018) for a comparison between ex-ante and ex-post screening systems in the Belgian context.

[5] The data on the distance between all pairs of Italian towns is available at https://www.istat.it/it/archivio/157423

Table 1 Descriptive statistics on travel time in minutes

	d	d_{univ}	d_{field}
Italy	69.2	14.0	23.7
North-East			
Average	*47.7*	*12.5*	*20.7*
Emilia Romagna	39.9	11.0	16.2
Friuli Venezia Giulia	55.5	12.1	20.8
Trentino Alto Adige	75.2	15.3	27.3
Veneto	48.5	13.4	23.5
North-West			
Average	*33.7*	*10.9*	*20.2*
Liguria	43.7	7.9	21.4
Lombardy	31.4	10.2	18.5
Piedmont	34.2	13.5	23.5
Aosta Valley	75.4	13.0	35.2
Central			
Average	*42.8*	*11.2*	*20.5*
Lazio	30.7	6.1	13.9
Marche	76.1	12.7	29.5
Tuscany	44.9	18.9	27.2
Umbria	53.3	13.7	22.0
Southern			
Average	*101.8*	*17.9*	*28.0*
Abruzzo	97.8	17.3	26.6
Basilicata	213.5	30.8	53.9
Calabria	170.7	33.1	46.7
Campania	41.7	13.3	18.6
Molise	134.9	12.5	40.5
Apulia	152.0	17.4	30.4
Islands			
Average	*170.7*	*20.6*	*33.9*
Sardinia	78.3	24.6	42.5
Sicily	198.5	19.3	31.3

The table reports, for each region and macro area of residence, the average distance between students' cities of residence and the chosen university (d), the nearest university (d_{univ}), and the nearest university that provides a program in the chosen field (d_{field})

unidirectional flow of students from southern to northern regions. This considerable difference among macro regions is not mirrored by the data regarding the distance to the nearest university, where the average distance varies between a minimum of 10.9 min in the North-West and a maximum of 20.6 min in the Islands. Indeed, we can see that, although students in southern regions and Islands need to travel more to reach the nearest university with respect to the others, this difference is minimal. Nevertheless, if we look at the differences

between regions, we have a different picture: the values for minimum distances are more heterogeneous, ranging from a minimum of 6.1 min in Lazio to a maximum of 33.1 min in Calabria if we consider all fields of study, and from 13.9 min in Lazio to 53.9 min in Basilicata while considering only the chosen field.

These data highlight the presence of differences in the tertiary education supply across the country, with students in some regions that need to travel more to reach the nearest university and, therefore, may have a different perception of which university has to be considered as *local* or *non-local* as they have different habits to travel for studying reasons. To account for this element and to avoid inconsistent assumptions that rely only on administrative borders or deterministic thresholds, we classify students' choices depending on two values: $d_{i,univ}$, given by the distance between student i's town of residence and the nearest university, and $d_{i,field}$, given by the distance between student i's town and the nearest university that provides a degree program in the chosen field of study. Moreover, since many Italian universities are located close to each other but in different cities, also universities located close to the nearest institution may be perceived as *local*. To account for this element, we apply a non-deterministic approach to define students' local areas. Our strategy consists in increasing both thresholds by a random amount of time $\delta \in [0;30]$ minutes and estimating a separate model for each value of δ. As we will show in Sect. 4, we generate $m = 5$ random values of δ_m to consider the uncertainty in the thresholds' definition, and then we combine the results of the analysis carried out using each value of δ_m by using Rubin's rule (Rubin 1987). It is worth highlighting that (i) since the distance is computed from cities the value of 0 can be observed only for universities in the same city, and that (ii) results do not change substantially by considering alternative and even more extreme intervals for δ_m (Porcu et al. 2021).

Following this strategy, we define three categories of university choices: *local*, *forced non-local*, and *free non-local*. Universities are considered *local* if placed closer than $d_{i,univ} + \delta_m$ minutes of travel from the student's town of residence, while they are considered *non-local* if placed farther than $d_{i,univ} + \delta_m$ minutes. For example, with a value of $\delta = 0$, the local area of all the students that reside in Milan comprises all the universities in the city of Milan. Instead, suppose a value of $\delta = 20$ is considered. In that case, the local area of students residing in Milan also includes other cities such as Sesto San Giovanni and Desio where there are two separate branches of the University of Milano-Bicocca.

Moreover, to account for differences in universities' supply that depend on the chosen field of study, we distinguish between two types of *non-local* universities: *forced* and *free*. Universities are considered *forced non-local* if they are located closer than $d_{i,field} + \delta_m$ minutes of travel from the student's town. Namely, a *forced non-local* university is the nearest university providing a program in a student's field of study. Finally, *free non-local* universities are those located farther than both thresholds.

3.3 Secondary schools

MOBYSU.IT database allows us to precisely identify the high school attended by each student. It provides information on the town hosting the school, the specific curriculum provided, and whether it is a private or a public funded one. We score 20,164 high schools.

Each observed high school provides several types of programs with significant differences in disciplinary contents. Table 9 in the Appendix lists the 13 high school curricula we consider in the analysis. Each curriculum is usually established at the national level, with a set of compulsory subjects that are taught in all the programs (mathematics, sciences, literature, history, one or two foreign languages, and gym classes) and a set of subjects that depend on the type of curriculum attended. Besides the differences in specializations, the curricula also differ in the amount of time allocated to each subject and expected outcomes. Indeed, following Contini et al. (2017), we can group these curricula into three general categories of institutes: the *liceo*, technical schools, and vocational schools. *Liceo* schools are generally oriented to provide students with a solid academic background and present four specializations: humanities, sciences, languages, and arts. Technical institutes provide students with a specialization in a particular field (e.g., accounting, surveyor, industrial) and general education. Vocational institutes are even more oriented to training students to enter the workforce with lab and professional activities programs.

These differences in programs and curricula are likely to impact also students' choices regarding *local* and *non-local* universities. Indeed, students with a more specialized background may prefer a particular university or, for example, students from a *liceo* are more likely to come from advantaged socioeconomic families, which, on average, can provide more financial resources to support their studies in tertiary education (Barone and Assirelli 2020; Impicciatore and Tosi 2019).

We account for these differences in two ways. First, we consider the type of high school attended among the predictors of the model. Second, as explained in Sect. 4, we apply a Cross-Classified Multinomial Logit (CCMNL) by defining the second level of clustering as the interaction between the high school and the curricula offered.

3.4 Students' characteristics

For each observed student, MOBYSU.IT provides information on several individual characteristics affecting students' university choices. In particular, we observe students' gender, age, the town of residence, school final grade, the attended high school, and the high school curriculum. Moreover, since we observe the years of diploma and enrollment, we can calculate the years between high school graduation and enrollment (indicated as *Late Enrollment*) and identify the students who obtained their high school diploma after turning 19 (indicated as *Irregular*). To account for students' commuting experience at high

school, we also computed the *High school mover* indicator that takes value 1 if the high school attended was not in the student's city of residence. Table 2 shows the descriptive statistics on students' characteristics by macro region of residence.

From Table 2, we can see that 64.9% of university students in our data reside in central or northern regions, while 35.1% in southern regions and Islands. As for students' characteristics, most students are females in all the macro regions and the average diploma grade is similar among areas, with lower average values in the North-West regions. Students generally enroll at university the same year when they get their high school degree. Indeed, the average number of years between high school graduation and enrollment is 0.12 at the national level. Most of the enrolled students were regular in the high school time schedule, with 13.2% who got their high school diploma with at least one year of delay. This share is lower in southern regions. Moreover, 50.9% of students in the country attended a high school in a different city than their city of residence. This share is higher in northern regions. The second panel of Table 2 reports information on the distribution of students among high school institutes. As expected, 70.5% of enrolled students came from a *liceo*, while only 21.7% came from a technical school. Students from vocational schools are the residual category, with an average share in

Table 2 Descriptive statistics on students' characteristics by macro area of residence

	Italy	Islands	South	Centre	North East	North West
Share of enrolled students		0.104	0.248	0.208	0.186	0.255
Students' characteristics						
Females	0.546	0.550	0.555	0.537	0.544	0.546
Diploma grade [60;100]	80.94	81.92	83.29	80.56	80.09	79.18
Late Enrollment (years)	0.125	0.161	0.119	0.115	0.134	0.117
Irregular	0.132	0.127	0.072	0.137	0.155	0.171
High School Mover	0.509	0.419	0.478	0.400	0.585	0.611
High School Institute						
Lyceum	0.705	0.733	0.756	0.731	0.630	0.679
Technical	0.217	0.202	0.182	0.193	0.278	0.233
Vocational	0.078	0.065	0.062	0.076	0.093	0.088
Enrollment year						
2016	0.175	0.171	0.178	0.179	0.170	0.173
2017	0.191	0.188	0.191	0.192	0.190	0.192
2018	0.196	0.194	0.198	0.196	0.194	0.194
2019	0.206	0.207	0.209	0.201	0.207	0.205
2020	0.233	0.240	0.225	0.233	0.238	0.235

The first panel shows the shares of enrolled students for each macro area. The second panel shows the shares of female students, the average diploma grade, the average number of years between diploma and enrollment, the shares of students that obtained their diploma after turning 19, and the one of students that were movers during their high school. The third panel shows the shares of students by the type of high school curriculum attended. The lastpanel shows the shares of students with respect to the year of enrollment

the country of 7.8%. The last panel shows that the number of enrolled students has increased in the last five years, with similar patterns in all the macro regions.

Finally, to control for the economic conditions of students' area of residence, we have collected from ISTAT the data regarding the provincial unemployment rate and the regional GDP per capita. Indeed, students in less economically developed areas, or regions with higher unemployment rates, may be more likely to migrate to a different region to seek better employment opportunities (see for example Cattaneo et al. 2017b; Giambona et al. 2017). Moreover, to account for differences in students' preferences that may be related to regional characteristics, we estimate a separate model for each macro area of residence. This strategy allows us to model the heterogeneity in students' preferences that depends on differences related to their macro area of residence without including also a full set of interactions terms between each predictor and the macro area fixed effects. Moreover, to account for geographical differences within each macro area we add the set of provincial fixed effects. This strategy allows us to control for all the time-invariant attributes that vary among provinces within the same macro area. Thanks to this strategy we aim to disentangle the high school effect on students' mobility choices from the contextual effects that depend on regional and provincial differences.

4 Empirical analysis

To estimate the influence of the attended high school on students' choices on whether to enroll in a *local* or *non-local* university, we specify a CCMNL model with three levels. Students (level-1 units) are classified according to the interaction between the attended high school and the curriculum (level-2) and the chosen field of study (level-3). Specifically, we observe 1,021,365 students cross-classified according to 126,138 high school/curricula pairs and 10 fields of study. This specification allows us to estimate the effect of high schools on students' choices by accounting for the variability in students' propensity to choose a *non-local* university that depends on differences among fields of study and students' characteristics. Moreover, as explained in Sect. 3.4, we account for the heterogeneity in students' preferences that depend on their macro area of origin by estimating 5 models, one for each considered area (North-East, North-West, Centre, South, Islands).[6]

Let $\pi_{ijkg} = P(Y_i = j | X, k, g)$ be the probability of student i to choose category j given the set of observable characteristics X, the high school/curriculum k, and the chosen field of study g. As explained in Sect. 3.2, students' university choices are classified into three categories depending on: the distance traveled by the student d_i, the minimum distance between students' town of residence and the nearest university $d_{i,univ}$, and the minimum distance needed to reach the nearest university which provides a program in the chosen field of study $d_{i,field}$.

---

[6] To keep the notation tractable, we do not include the additional set of subscripts related to students' macro areas of residence.

As said above, to classify universities as *local* or *non-local* with respect to each student's local area, a random amount of time δ_m has been added to these distances. In particular, j can take three values:

- $Y_i = 0$, i chooses a *free non-local* university, i.e. $d_i > d_{i,field} + \delta_m$
- $Y_i = 1$, i chooses a *forced non-local* university, i.e. $d_{i,univ} < d_i \leq d_{i,field} + \delta_m$
- $Y_i = 2$, i chooses a *local* university, i.e. $d_i \leq d_{i,univ} + \delta_m$

This classification of students' choices allows us to differentiate between *local* and *non-local* universities by accounting for the actual supply of universities in students' local area and the chosen field of study. Indeed, the minimum distances are computed for each student by considering her/his town of residence. Nonetheless, we distinguish between students who move to attend a program that is not available in their local area from the ones that migrate even when the nearest university provides a program in their field of study, introducing two categories for classifying non-local choices: *forced* and *free non-local*.

Given this setting, and considering *free non-local* university (i.e. $Y_i = 0$) as our baseline, the probability of student i to choose between *local* (i.e. $Y_i = 2$) and *forced non-local* (i.e. $Y_i = 1$) is estimated by using a multinomial logistic function:

$$\log\left(\frac{\pi_{ijkg}^m}{\pi_{i2kg}^m}\right) = X'\beta_j^m + u_{jk}^m + v_{jg}^m \text{ with } j = 0, 1 \text{ and } m = 1, \ldots, 5 \tag{1}$$

where X is the set of variables used in estimation, $u_{jk} \sim \mathcal{N}(0, \Omega_u)$ is the random intercept that captures the between school/curriculum (k) variability, as it is shared by students nested in the same high school/curriculum. Meanwhile, $v_{jg} \sim \mathcal{N}(0, \Omega_v)$ is the random intercept that captures the between fields of study variability which is shared by students who enrolled in the same field of study. The superscript m indicates that several models are estimated considering different values of the random term δ_m. X contains several variables that account for students' characteristics, their high school background and the characteristics of their residence area. In particular, we apply the empirical approach by estimating Eq. 1 for each macro area of residence separately and including the set of provincial fixed effects to account for geographical differences within each macro area in students' propensity to choose *local* universities. In Table 3 we report the definitions of each regressor used in estimation, along with the average value observed and the standard deviation.

Concerning the random parameters, the joint introduction in the predictors of the two random terms, u_{jk} and v_{jg}, allowed us to capture divergences in the preferences for *local* and *non-local* universities over and above the individual characteristics and the preferences for different fields of study. Therefore, the set of parameters u_{jk} captures the residual heterogeneity in students' mobility choices between students attending different high school/curriculum pairs after controlling for the differences in students' preferences that depend on their characteristics and the chosen field of study. Thus, given this specification, the posterior estimates of u_{jk} can be used as

Table 3 Definitions and descriptive statistics of variables used in estimation

Variable	Definition		
Continous variables:		Mean	Std dev
Diploma grade	High school diploma grade [60;101]	80.938	11.979
Late Enrollment	Number of years between the high school diploma and the enrollment	0.125	0.439
Regional GDP	Regional added value per capita in thousands of euros	25.542	6.983
Unemployment	Provincial unemployment rate	11.798	6.201
Categorial variables:		Proportion	
Female	1 if the student is female	0.546	
Irregular	1 if the student obtained the diploma after turning 19	0.132	
Technical inst.	1 if the student attended a technical institute	0.217	
Vocational inst.	1 if the student attended a vocational institute	0.078	
School mover	1 if the high school attended was not in student's city of residence	0.509	
Enrolled in 2016	1 if the student enrolled in 2016	0.175	
Enrolled in 2017	1 if the student enrolled in 2017	0.191	
Enrolled in 2018	1 if the student enrolled in 2018	0.196	
Enrolled in 2019	1 if the student enrolled in 2019	0.206	
Enrolled in 2020	1 if the student enrolled in 2020	0.233	
Provincial fixed effects	Set of indicators variables that take value 1 if the student resides in the considered province		

Note: The first panel shows the sample mean and the standard deviation for each continuous variable used in estimation. The second panel shows the proportions for each indicator variable included in the estimation. See Table 10 in the Appendix for the distribution of students by region of residence

a measure of the influence of the high school environment on students' mobility choices.

Finally, to account for the uncertainty in the definition of students' local areas and the related classification of students' university choices as *local*, *free non-local*, and *forced non-local*, we estimate the model for each macro area of origin by considering $M = 5$ values for δ_m selected as random draws from the interval $\delta_m \in [0;30]$ minutes. The results are combined by using Rubin's rule (Rubin 1987) to obtain a single statement for each macro area of residence considered. Formally, denoting by $\theta^m = [\beta_j^m, u_{jk}^m, v_{jg}^m]$ the set of parameters obtained in each estimation m, the pooled estimate $\bar{\theta}$ is computed as the average value over the M estimates:

$$\bar{\theta} = M^{-1} \sum_{m=1}^{M} \hat{\theta}^m. \tag{2}$$

The standard error related to each parameter $\bar{\theta}$ is obtained as:

$$SE(\bar{\theta}) = \sqrt{W + B + \frac{B}{M}} \tag{3}$$

where W is the within estimates variance:

$$W = M^{-1} \sum_{m=1}^{M} SE(\hat{\theta}^m)^2 \tag{4}$$

and B is the between estimates variance:

$$B = \frac{\sum_{m=1}^{M}(\hat{\theta}^m - \bar{\theta})^2}{M - 1} \tag{5}$$

According to the steps described above, we can estimate the influence of secondary school on students' mobility choices by accounting for the uncertainty in the definition of the thresholds $d_{i,univ} + \delta_m$ and $d_{i,field} + \delta_m$.

The main advantage of adopting such an approach to deal with missing information is that the estimates of parameters explicitly rely on different values of the thresholds to define a university as *local* or *non-local*, and the standard errors specifically account of the related uncertainty.

5 Results

We begin by reporting the parameter estimates of the CCMNL, together with a discussion of each predictor's effect on students' choice probabilities. Then, we focus on the role of high schools by exploiting the posterior parameter estimates associated with the random intercepts u_{jk}. In particular, we show how the high schools' effect, defined as the value associated with each u_{jk}, varies among students and high schools.

5.1 Cross-classified multinomial logit results

Table 4 reports, for each macro area of residence, the parameter estimates obtained by pooling the 5 vectors of coefficients resulting from estimating the set of CCMNLs, one for each value of δ_m. All the parameters have been estimated with the `runmlwin` routine in `Stata` which allows to apply a Monte Carlo Markov Chain algorithm through the program `MLwiN` (Leckie and Charlton 2013).

Table 5 depicts the distribution of students in the three identified choice categories, which describe the preferences for *local* or *non-local* universities for the different values of δ_m. It is worth highlighting that the definition of the thresholds has a relevant influence on the classification of a university as *local*, *forced non-local* and *free non-local*. The difference in the rate of *free non-local* is about 15% between the two extreme values of δ_m. However, models' estimates for each value of δ_m are similar in terms of their signs and magnitude.[7] This element suggests

---

[7] The results regarding each CCMNL estimated for each value of δ_m are available from the authors upon request.

Table 4 Pooled CCMNL (Baseline = Free Non-Local university)

	North-East		North-West		Centre		South		Islands	
	Local	Forced Non-local	Local	Forced Non-local	Local	Forced Non-local	Local	Forced Non-local	Local	Forced Non-local
Constant	−0.2325	−3.6029***	5.5838***	5.8139***	−0.2739	−4.5906***	−4.2353**	−2.0582**	−0.2791	4.3915***
	(0.4040)	(0.4203)	(1.1464)	(2.1105)	(0.4909)	(0.6461)	(1.7251)	(0.8647)	(0.6413)	(0.7913)
Female	−0.2018***	−0.0880***	−0.0745***	−0.0601***	−0.1131***	−0.0486*	0.0052	0.0368	0.0490	0.0825**
	(0.0249)	(0.0273)	(0.0154)	(0.0203)	(0.0216)	(0.0265)	(0.0522)	(0.0715)	(0.0360)	(0.0355)
Diploma grade	−0.0051***	−0.0004	−0.0066***	−0.0022*	−0.0160***	−0.0106***	−0.0138***	−0.0087***	−0.0167***	−0.0053***
	(0.0008)	(0.0013)	(0.0010)	(0.0012)	(0.0019)	(0.0011)	(0.0019)	(0.0027)	(0.0010)	(0.0015)
Irregular	−0.0486	−0.1208***	−0.0018	−0.0847***	−0.0542*	0.0023	−0.0587	−0.0997**	0.0364	0.0730*
	(0.0390)	(0.0275)	(0.0626)	(0.0226)	(0.0294)	(0.0343)	(0.0374)	(0.0413)	(0.0394)	(0.0444)
Technical inst.	0.4438***	0.4414***	0.3882***	0.3256***	0.6019***	0.6369***	0.5672***	0.4044***	0.8967***	0.5126***
	(0.0827)	(0.0398)	(0.1155)	(0.0959)	(0.1565)	(0.0712)	(0.0645)	(0.0621)	(0.0567)	(0.0771)
Vocational inst.	0.4322***	0.2576***	0.3474***	0.2121*	0.4292***	0.5756***	0.4401***	0.3747***	0.9892***	0.6767***
	(0.0390)	(0.0763)	(0.0409)	(0.1159)	(0.1373)	(0.0741)	(0.0703)	(0.0499)	(0.0770)	(0.0779)
School mover	0.2153	0.4090	0.2291	0.4146	0.0185	0.3281	0.1714	0.1636*	0.2075***	0.3865***
	(0.3143)	(0.2901)	(0.5650)	(0.3316)	(0.3214)	(0.2362)	(0.1607)	(0.0939)	(0.0524)	(0.0712)
Late Enrollment	0.0176	0.0237	−0.0548	−0.0629***	−0.0908***	−0.0905**	−0.0215	−0.0459	−0.1231***	−0.0601*
	(0.0221)	(0.0307)	(0.0372)	(0.0240)	(0.0275)	(0.0354)	(0.0413)	(0.0419)	(0.0289)	(0.0349)
Unemployment	0.0024	0.0138	−0.0507	−0.0395	−0.0069	−0.0071	0.0074	0.0078	0.0314***	0.0226*
	(0.0198)	(0.0194)	(0.0331)	(0.0279)	(0.0164)	(0.0091)	(0.0103)	(0.0070)	(0.0090)	(0.0128)
Regional GDP	−0.0662***	0.0549**	−0.1582***	−0.2624***	0.0096	0.1463***	0.2329***	0.0269	−0.2076***	−0.4284***
	(0.0201)	(0.0218)	(0.0095)	(0.0323)	(0.0237)	(0.0174)	(0.0850)	(0.0272)	(0.0354)	(0.0461)
Enrolled in 2017	−0.0827**	−0.1544**	0.0230	0.1468***	−0.0095	−0.1259**	−0.0620	−0.0253	0.0377	0.0348
	(0.0386)	(0.0644)	(0.0381)	(0.0397)	(0.0465)	(0.0530)	(0.0553)	(0.0935)	(0.0743)	(0.0888)

Table 4 (continued)

	North-East		North-West		Centre		South		Islands	
		Forced		Forced		Forced		Forced		Forced
	Local	Non-local	Local	Non-local	Local	Non-local	Local	Non-local	Local	Non-local
Enrolled in 2018	−0.1187**	−0.2033***	0.0483	0.2833***	0.0963*	−0.1047*	−0.0788	0.0130	0.2775***	0.2726***
	(0.0524)	(0.0711)	(0.0574)	(0.0855)	(0.0580)	(0.0546)	(0.0922)	(0.0589)	(0.0844)	(0.0886)
Enrolled in 2019	−0.0520	−0.2791***	0.1239*	0.4373***	0.0517	−0.1914***	−0.0957	0.0673	0.5039***	0.4611***
	(0.0557)	(0.0886)	(0.0732)	(0.1208)	(0.0614)	(0.0642)	(0.1185)	(0.0701)	(0.0866)	(0.0837)
Enrolled in 2020	−0.1818***	−0.3529***	0.0560	0.5136***	−0.0567	−0.3238***	−0.2254	0.0699	0.5016***	0.6582***
	(0.0676)	(0.0943)	(0.0840)	(0.1269)	(0.0798)	(0.0692)	(0.1643)	(0.0908)	(0.0987)	(0.0975)
Provincial F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School × Curr.:										
Variance	1.4189***	1.1550***	2.3539***	1.3644***	3.2374***	1.7862***	2.8323***	1.4749***	4.1412***	2.2358***
	(0.1101)	(0.2613)	(0.3816)	(0.1410)	(0.4387)	(0.2193)	(0.6612)	(0.1523)	(0.1669)	(0.1757)
Covariance		0.1321**		0.3870**		0.0040		−0.0821		0.6457***
		(0.0541)		(0.1575)		(0.1347)		(0.1223)		(0.1656)
ISCED fields:										
Variance	0.4369	2.2842	4.0125	10.2181	0.7768	2.6812	1.5146	3.4038	0.3232*	0.7628*
	(0.3769)	(2.6514)	(2.6711)	(8.5641)	(0.6542)	(2.0183)	(1.9531)	(2.7790)	(0.1932)	(0.4345)
Covariance		0.4883		1.0918		1.0110		1.9493		0.3491
		(0.9324)		(3.9478)		(1.0342)		(2.1261)		(0.2452)
Observations	189,582	189,582	260,396	260,396	212,581	212,581	253,050	253,050	105,756	105,756

Note: Parameters and standard errors (in parenthesis) are computed according to Sect. 4. For each macro area the first column refers to the contrast between *Local* and *Free non-local* universities, while the second refers to the one between *Forced non-local* vs *Free non-local* universities. The first panel shows the set of estimated β, the second and third panels show the variances and covariances of the random intercepts distributions. ***, **, * indicate, respectively, that the coefficient is significant at the 1%, 5% or 10% level

Table 5 Shares of students in each choice category

Estimate	δ_m (min)	Local	Forced	Free
1	10.5	0.48	0.18	0.33
2	8.0	0.45	0.20	0.35
3	4.1	0.40	0.23	0.37
4	0.9	0.37	0.23	0.40
5	26.1	0.64	0.11	0.25

The table reports, for each model estimation from 1 to 5, the share of students in each choice category for each value of δ_m

that final results do not depend on the specific thresholds used to define students' local areas. Indeed, using a non deterministic approach, considering different values for δ_m, ensures that the parameter estimates of the CCMNL model take into account the uncertainty in the definition of students' local areas.

Turning on the pooled estimates presented in Table 4, we focus on interpreting the main results related to students' and local areas' characteristics in this subsection. In the following subsection, we will address the results regarding the posterior estimates of random intercepts.

The first element that it arises in Table 4 is related to the variation in the magnitude and signs of some coefficients when considering different macro areas. Moreover, as it will be clear from the discussion on the average marginal effects in Table 6, we can notice that there is a relevant difference in the relative magnitude and signs associated with the constants. This element suggests an heterogeneity in the average probability to choose one of the considered alternative that depends on differences among macro areas.

As for other students' characteristics, we can see that females are more likely to choose a *free non-local* university than males when residing in northern or central regions of the country, with a more substantial effect in the contrast *local* vs *free non-local*. Instead, if we look at students from the South and Islands the coefficients are positive, even though with a lower relative magnitude with respect to the other characteristics (i.e. the type of high school curricula or the high school mover status). This result may be related, among others, to two elements. First, the presence of some cultural traits that may incentive families in the South to invest more in their sons' education than their daughters' one (see, for example, the discussion in Ballarino et al. 2022). Second, the uneven distribution in the country of some gender oriented degree programs and the heterogeneity in their quality (e.g. most important polytechnics are located in the North) may motivate males to leave their region of residence more often than females (D'Agostino et al. 2019a; Gibbons and Vignoles 2012).

Moreover, students who have obtained their high school degree in technical or vocational institutes and those who had to commute to reach their high school have a higher probability of choosing a *local* or *forced non-local* alternative. These results are similar across the macro areas and confirmed by analysing the average marginal effects reported in Table 6.

Looking at the economic characteristics of students' areas of origin, once we control for the differences between provinces, the regional GDP has a heterogeneous effect based on the considered macro area. It negatively influences the probability of selecting a *local* university in northern regions and the Islands, while the effect is positive for students in the Centre (despite negligible) and in southern regions. Concerning *forced non-local* universities, we can see that the GDP has a negative

Table 6 Average marginal effects of students' choices main determinants

	Local		Forced non-local		Free non-local	
	π_0 (%)	Δ	π_1 (%)	Δ	π_2 (%)	Δ
Average student	47.2		17.0		35.7	
Centre	54.0		18.5		27.5	
North West	47.0	−7.01	24.2	5.70	28.8	1.31
North East	37.1	−16.93	13.9	−4.56	48.9	21.49
South	45.8	−8.24	13.7	−4.79	40.5	13.03
Islands	55.5	1.42	10.0	−8.49	34.5	7.08
Male	47.7		17.1		35.2	
Female	46.8	−0.96	17.0	−0.01	36.2	0.97
Liceo	45.7		16.3		38.0	
Vocational Institute	50.3	4.60	18.3	2.00	31.4	−6.61
Technical Institute	50.9	5.20	18.8	2.48	30.3	−7.68
High school mover = 0	46.7		15.4		37.9	
High school mover = 1	47.7	0.99	18.2	2.80	34.1	−3.79
Irregular = 0	47.2		17.1		35.7	
Irregular = 1	47.0	−0.19	16.6	−0.56	36.4	0.75
Average grade	47.2		17.0		35.7	
Diploma grade (12)	45.8	−1.46	16.9	−0.15	37.4	1.61
Years of late enrolment = 0	47.3		17.1		35.7	
Years of late enrolment = 1	46.9	−0.34	16.8	−0.31	36.3	0.66

The table reports the average predicted choice probabilities (π_j) and the average marginal effects (Δ) for each profile of student. Each probability is estimated by using the coefficients obtained from the model estimated considering students' macro area of residence. For example, the average probabilities associated with the *Liceo* indicator are computed by fixing the *Liceo* indicator equal to one and holding the other variables at their observed values. The second column reports the average marginal effects computed as the difference between the probabilities related to the considered profile and its baseline. For example, the average marginal effect associated with the female indicator for the *local* alternative (π_0) is computed as the difference between the average probability for males (47.7%) with the one for females (46.8%). The average marginal effects are computed for continuous variables by increasing their values by one standard deviation (reported in parenthesis). See also Long and Mustillo (2021) for the computation of average marginal effects as average discrete changes

and relevant effect in terms of magnitude when considering students coming from the North-West and in the Islands while it is positive for students that reside in the North-East, in the Centre, or the South. In contrast, the provincial unemployment rate has a negligible effect in most of the areas considered. However, it is important to note that the interpretation of these coefficients is not trivial. Indeed, the inclusion of provincial fixed effects may absorb a relevant part of the effect related to GDP and unemployment. Therefore, the inclusion of these variables in the model has the purpose to account for the variation in geographical economic conditions within the country rather than to have an estimation of the effect of these characteristics.

In general, we can see that all the coefficients associated to students' characteristics have the same sign in influencing the probability of both outcomes in all the macro areas considered. This element suggests that *local* and *forced non-local* alternatives are considered similar from students' perspectives. These alternatives attract students who prefer universities located within or near their local area, depending on the availability of a program in their field of study.

To better understand these results, we reported in Table 6 the average marginal effects on students' choice probabilities of the main predictors included in Table 4 and the differences in probabilities associated with students' macro areas of residence. In particular, for each alternative considered, we report two values: the predicted choice probability associated with each profile and the difference in choice probabilities between each profile and its baseline. Following Long and Mustillo (2021), we computed the average marginal effects as average discrete changes. For example, the choice probabilities associated with the *vocational institute* indicator are computed by predicting the individual probabilities assuming that all the students have attended this high school curriculum and holding all the other characteristics to their observed values. These probabilities are computed by using the estimates associated with students' macro areas of residence. Then, the average marginal effect associated with the indicator is computed by taking the difference between the choice probabilities related to the *liceo* profile and those associated with the *vocational institutes*. In this case, our results indicate that, *ceteris paribus*, the probability of choosing a *free non-local* university for students with a vocational or technical diploma is, respectively, 6.61 and 7.68 percentage points lower than the one for those who attended a *liceo*. This result can be related to the fact that students from the *liceo* are usually more oriented toward an academic career and, therefore, may be more interested in university characteristics besides the distance from their towns, such as quality of research and job opportunities. These elements may encourage students from *liceo* high schools to leave their local area to reach a university that satisfies their preferences better. With respect to the *high school mover* status, we can see that students who travelled to reach their high school in the past prefer to enroll in a *forced non-local* university more than the other students. This indicates that these students are willing to travel more to reach the first university providing a program in their preferred field of study. Indeed, most of these students do not have a university in their town of residence and, therefore, must commute to reach the nearest university. Thus, they may perceive as similar, in terms of travel

time, *local* and *forced non-local* universities, and choose the second because it provides a program in their field of study.

As for students' macro areas of residence, we can see that this is one of the most important determinants of students' choices: changes in this variable are associated with the most relevant changes in students' choice probabilities. From the first column of Table 6 we can notice that the majority of students in central regions and in the Islands are enrolled in *local* universities. Ceteris paribus, students from the North-East are those with the lowest probability to choose a *local* university and are more likely to choose a *free non-local* alternative. In general, the reduction in the probability to choose a *local* alternative is associated with an increase in the one related to the *free non-local* alternative. The only exception is given by students in the North-West who prefer to enroll in *forced non-local* alternatives rather than *free non-local*. These results confirm that, from students' perspective, *local* and *forced non-local* alternatives are perceived as similar. Moreover, the fact that students residing in the North-West are less mobile provides evidence of the role that these regions have in attracting students from other areas of the country. These differences among macro areas are also related to the existence of a strong and persistent flow of students who migrate from the Islands and southern regions to the Centre-North of the country. As for students' past performances and careers, we can see that those with higher diploma grades prefer the *free non-local* alternative over the *local* alternative. A standard deviation increase in the diploma grade (12 points) is associated with an increase of 1.61 percentage points in *free non-local* choice probabilities and a reduction of 1.46 points in *local* choice probabilities. In contrast, the effect on the *forced non-local* alternative is lower (−0.15). Moreover, we can note that, once controlling for other students' determinants, their irregular status and years of delay in enrollment do not have a strong impact on location choices.

Finally, concerning students' sex, we have that, on average, female students are 0.97% more likely to choose a *free non-local* university with respect to males. This effect is almost totally driven by a reduction in *local* choice probabilities. It is worth mentioning that findings on this aspect do not agree in indicating a higher probability of one gender to be in mobility with respect to another (see, for example, Cattaneo et al. 2017a; Columbu et al. 2021; D'Agostino et al. 2019a), but results seem to provide opposite evidence. Indeed, as shown with respect to the results in Table 4, the differences between females and males are not constant across the country, with females in southern regions and in the Islands that are more prone to choose *local* alternatives. Moreover, D'Agostino et al. (2019a) show a prevalence in the rate of movers of males in STEM students in Italy between 2008 and 2014. These results have to be considered taking into account the high variability in the students' propensity to be in mobility between disciplinary fields and that the choice of the disciplinary field is indeed gender-oriented and affected by family socioeconomic conditions. Moreover, findings suggest that female students have a higher propensity to be in mobility when they come from a high socioeconomic background (see, for example, Gibbons and Vignoles, 2012).

5.2 The high school effect

In this section, we address the interpretation of the results concerning the set of estimated random intercepts by focusing mainly on the results associated with the high school/curriculum level parameters u_{jk}.

The bottom panel of Table 4 presents the information related to the set of random intercepts $u_{jk} \sim \mathcal{N}(0, \Omega_u)$ and $v_{jg} \sim \mathcal{N}(0, \Omega_v)$ for each macro area of residence. In particular, for each level, we have estimated the variances and the covariances that help to recover the variance-covariance matrices Ω_u and Ω_v. The results show that the variance estimated at the high school/curriculum level for contrast *local* vs *free non-local* is higher than the one estimated for the field of study level in all the models, with the exception of the one estimated for students residing in the North-West. This suggests that, in general, the differences among high school/curricula pairs (level k units) explain much of the variability in students' choices for *local* or *free non-local* universities that is not explained by the other predictors, especially if we consider students in the Centre or in the Islands where the variances in the first level are more than three times the ones estimated for the second level. Instead, if we consider the contrast between *forced non-local* and *free non-local* we can see that the variances estimated for the second level (fields of study) are higher than the ones estimated at the high school/curricula level in all the macro areas with the exception of the Islands. This result indicates that, as expected, the role of the field of study is more important when considering the choice to enroll in *forced non-local* universities.

To inform on the relative importance of each level of the cross-classification, Table 7 reports the variance partition coefficients (VPC) for each estimated model. This measure can be interpreted as the proportion of the unexplained variability in students' choices at each level of the multilevel structure (Leckie 2013; Snijders and Bosker 2011).[8]

The VPCs show that the differences in students' high school backgrounds play a crucial role in their decision process that depends on students' macro areas of residence. Indeed, the share of unexplained variation in students' choices for the contrast *local* vs *free non-local* that is explained by the variability at the high school/curricula level ranges from 24.38% in the North-West to the 53.41% in the Islands where the high school effect appears to be more important. As expected, the importance of high schools is lower when considering the contrast *forced* vs *free non-local* where the differences among fields of study have a prominent role. In this case, the VPCs associated with the high schools range from 9.17% in the North-West to 35.55% in the Islands. Therefore, these results indicate that the high school effect is weaker in northern regions with respect to the rest of the country.

Therefore, the results show that, once we account for covariates, the high school maintains a relevant role in affecting students' mobility choices, particularly if we consider the choice regarding the *local* university and students that reside in central or southern regions and the Islands.

---

[8] For example, suppose we indicate with σ_k^2 the variance for level k parameters and with σ_g^2 the variance for level g parameters. In that case, the VPC associated with level k parameters is given by: $\sigma_k^2 / (\sigma_g^2 + \sigma_v^2 + \pi^2/3)$.

Table 7 Variance partition coefficient

	Free non-local vs:	
	Local (%)	Forced (%)
North-West		
ISCED field of study	41.55	68.71
School × Curriculum	24.38	9.17
Individual level	34.07	22.12
North-East		
ISCED field of study	8.49	33.95
School × Curriculum	27.58	17.16
Individual level	63.93	48.89
Centre		
ISCED field of study	10.63	34.56
School × Curriculum	44.32	23.03
Individual level	45.04	42.41
South		
ISCED field of study	19.83	41.67
School × Curriculum	37.09	18.06
Individual level	43.08	40.27
Islands		
ISCED field of study	4.17	12.13
School × Curriculum	53.41	35.55
Individual level	42.43	52.32

Variance Partition Coefficients measure the proportion of unexplained variation in students' choices that lies at each level of the CCMNL. Each panel shows the results for each macro area of residence

Another valuable piece of information in Table 4 is the estimated covariance between the random intercepts in the two levels considered. As we can see, the estimated covariance at the high school/curricula level is very small when considering students residing in the South and the Centre of the country. These values become more relevant when considering the other areas but do not indicate the existence of a strong correlation between the two alternatives. This suggests that, from high schools' perspective, the two alternatives are unrelated: schools that advise students on *local* universities are not likely to suggest to their students to enroll also in a *forced non-local* university. This finding is somewhat surprising considering that the evidence presented in Sect. 5.1 indicates that students who choose a *local* university are similar to those that choose a *forced non-local* alternative.

A similar interpretation also applies to the estimated covariance at the field of study level, which results not significant with an estimated standard error that is often higher than its point estimate. This element suggests that the distributions of the random intercepts in the two contrasts are not correlated.

To highlight the importance of the high school in shaping students' choices, Figs. 1 and 2 depict the relationship between predicted choice probabilities and the estimated random intercepts at the high school/curriculum level. Negative values for
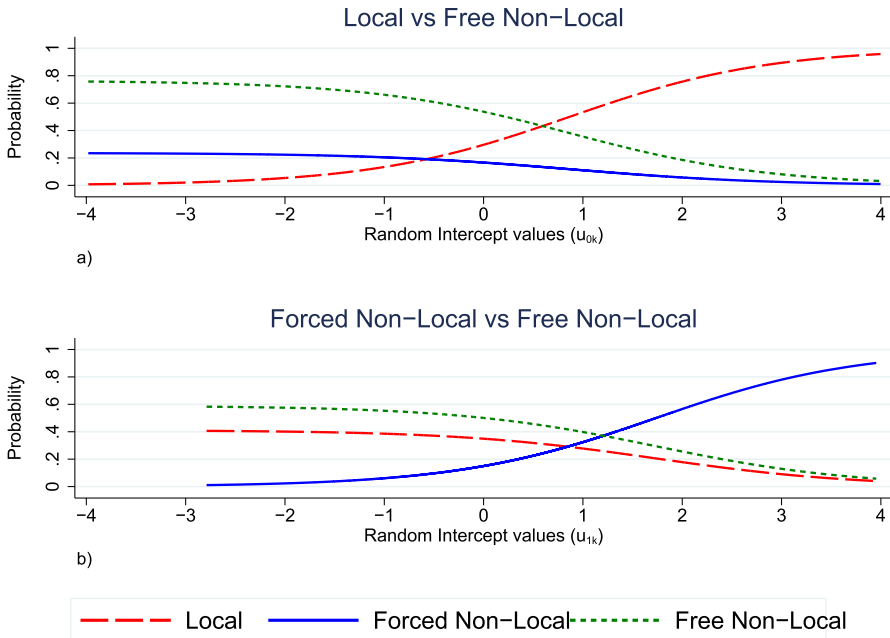
Fig. 1 Relationship between predicted choice probabilities and the high school effect. Note: The figure reports two plots depicting the relationship between the predicted choice probabilities and the estimated high school effect, defined as the set of high school/curriculum level random intercepts u_{jk}. The choice probabilities are computed by considering a student with average characteristics and by changing only the value of the random intercept considered according to the model estimated for students' macro areas of residence. The first row (plot a) shows the results for the *local* vs *free non-local* contrast (u_{0k}), while the second (plot b) regards the *forced* vs *free non-local* contrast (u_{1k})

the random intercepts can be interpreted as a propensity of high schools to refer students toward a *free non-local* alternative. In contrast, positive values indicate that schools have a positive effect in steering students toward *local* or *forced non-local* alternatives, depending on the contrast considered.

The plots in Fig. 1 are built up by computing students' choice probabilities according to the model estimated for their macro area of residence and then looking at how these probabilities vary when the high school effect changes for a student with average characteristics. The first plot considers the contrast *local* vs *free non-local* while the second considers the contrast *forced non-local* vs *free non-local*. From the first plot, we can notice that the high school has a key role in the comparison between *local* and *free non-local* universities. If we consider the average school (i.e. $u_{0k} = 0$), we can notice that the choice probabilities related to *local* and *forced non-local* alternatives are very similar between 20% and 30% with an advantage for the former. If we move from the average toward negative values, we can see that the probability of choosing *local* universities decreases with an increase in both *free non-local* and *forced non-local* choice probabilities. At the same time, if we consider schools that have a positive propensity to point students to *local* universities, we can
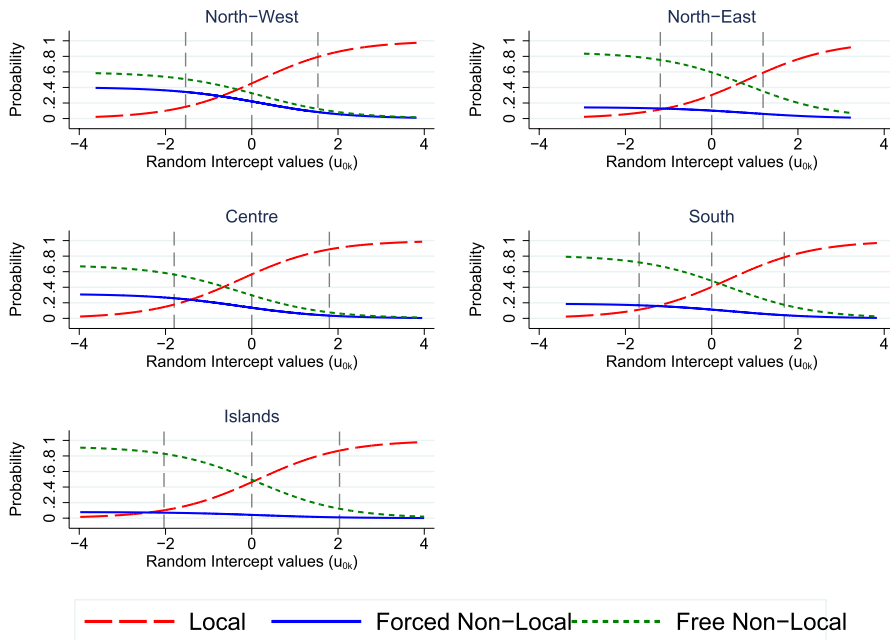
Fig. 2 The high school effect by macro area of residence, *Local* vs *Free non-local* contrast. Note: The figure reports the relationship between the predicted choice probabilities and the high school effect for the contrast *local* vs *free non-local* (u_{0k}). Predicted probabilities have been computed considering the average student in each macro area of residence. The dashed lines indicate the units with an average value of the random intercept equal to zero and those with a value of $+/-$ one standard deviation

see that the probability related to *local* alternatives increases by surpassing the one related to *free non-local* alternatives when the value for the random effect reaches a value close to 1, which is lower than the lowest standard deviation estimated for the distribution of the random effect u_{0k} (1.19 for North-East). This element suggests that schools have a relevant role in steering students toward *local* alternatives even when the estimated high school effect is not very different from the average.

From the second plot in Fig. 1 we can confirm that the high school has, in general, a less important role when considering the contrast *forced non-local* vs *free non-local*. Indeed, comparing the two plots, we can notice that the probability to choose the *forced non-local* alternative becomes prevalent with respect to both alternatives only when considering values of the random effect that are higher than 1. Namely, to make the student indifferent between the three alternatives the high school effect has to be stronger. This element is because the average predicted probability of choosing a *forced non-local* is much lower than the one related to the other alternatives (see Table 6).

Focusing on the high school effect with respect to the contrast between *local* and *free non-local* alternatives, Fig. 2 reports the relationship between the predicted choice probabilities and the high school effect considering the macro area of residence. Namely, each plot depicts how choice probabilities change as the school effect increases for the average student residing in a region of the considered macro area.

As we can see, the relationship between choice probabilities and the high school effect changes in each macro area considered. In line with the results discussed in Sect. 5.1, we can see that, if we consider the average school, the *free non-local* alternative is prevalent in the North-East while *local* alternatives are preferred by students in the Centre and in the North-West. In contrast, *free non-local* and *local* alternatives present similar choice probabilities in the Islands and southern regions. The results indicate that in these regions these two alternatives are perceived as similar and that even a small deviation from the average may have a relevant role in steering students toward *local* or *free non-local* alternatives.

Therefore, while in the Centre and North-West of the country there is an average tendency to choose a *local* alternative, in the other areas the high school effect has to be positive to keep students in their local area, especially if we consider students in the North-East. The difference between the North-East and the North-West can be also related to the heterogeneity in the university supply in these regions. Indeed, the North-West is characterized by the presence of important universities such as those hosted in Milan (e.g. Polytechnic of Milan, Milan-Bocconi). Moreover, this element helps to explain also the similarities between North-West regions and the Centre, where some important universities are hosted in Rome (e.g. Sapienza).

Nevertheless, these results indicate that high schools with a high than average propensity toward *local* universities may help in counteracting the asymmetric flow of students from southern to northern regions.

To better explore the role of high schools in the country, Fig. 3 shows the geographical distribution of the high school effects in Italian provinces. In particular, we have computed the average high school effect in each province by considering either students who have attended a *liceo* (on the left) or students who have attended a technical or vocational institute (on the right). In each map, the blue areas indicate that, on average, the high school effect is positive toward *local* universities, while the yellow-red areas indicate that, on average, the high schools are orienting more toward the *free non-local* alternative.

From the maps, we can highlight some important elements. First, as noted with respect to Fig. 2, the results concerning the *liceo* show that the high school effect is positive in the provinces that host very big universities like Milan, Turin, Florence, Naples, Rome and Bologna. Instead, if we look at Technical and Vocational high schools we estimate an average positive effect also in provinces that are more peripheral with respect to important universities. Moreover, we estimate a negative effect in Rome, Milan and Bologna. This result confirms the ones discussed with respect to marginal effects in Table 6: students in Technical and Vocational high schools prefer to stay in *local* or *forced non-local* universities. Considering Islands, we can see the attractive role of the two cities of Cagliari and Palermo. Another important element is the absence of a clear asymmetry between the northern and southern regions of the country. Indeed, we have areas with high schools with a strong propensity toward *local* alternatives in all the regions and geographical areas. These elements suggest that high schools may contribute to polarizing students' choices toward the universities that already attract higher numbers of students.

This evidence, together with the results reported in Fig. 2, helps to explain the existence of an asymmetric flow of students from the South and the Islands toward
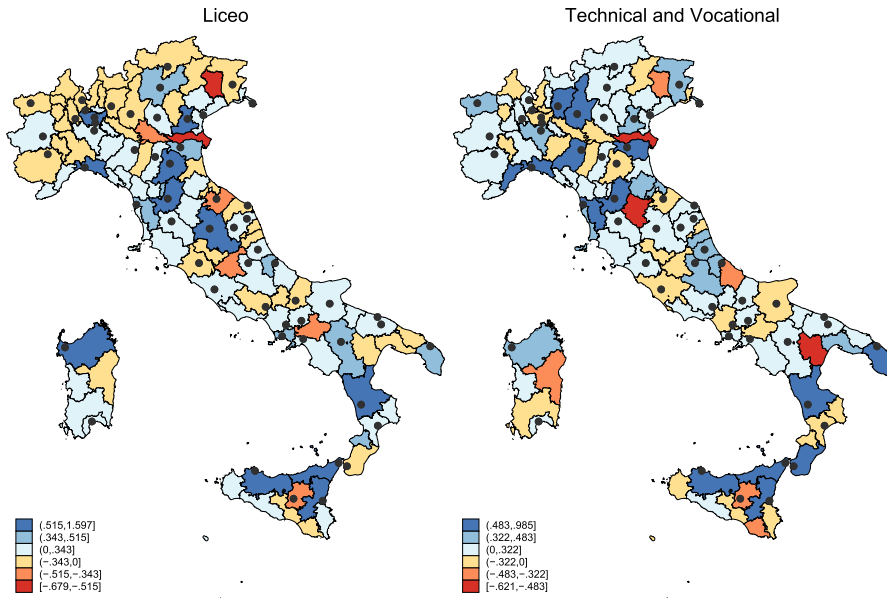
Fig. 3 The high school effect in Italy. Notes: The two maps consider, for each Italian province, the average value of the estimated high school effect for the contrast *local* vs *free non-local* considering the high school curriculum *liceo* (on the left) and the technical and vocational institutes (on the right). The high school effect is defined as the posterior estimate for the random intercepts u_{0k} obtained in the pooled CCMNL estimated for each macro area. The positive values (i.e. schools orienting toward *local* universities) are shown in blues, while negative values (i.e. schools orienting toward *free non-local* universities) are shown in yellow-red colors. For each observed university, the black dots indicate the town hosting the most important branch in terms of enrolled students

the northern and central regions and to provide some policy insights. Indeed, the results show that students in southern regions and the Islands prefer *free non-local* alternatives over *local* ones. However, the *local* alternative becomes prevalent in presence of a small positive high school effect toward them. However, as we can see from the map, this is the case only in the presence of big cities and universities that, in turn, are less common in these regions. However, this result suggests that an increase in high school policies that encourage students to choose *local* universities may play a role in countering the so-called North–South divide.

6 Conclusion

This work provides information related to the Italian context on how much of the divergences in students' mobility choices can be explained by differences in their high school backgrounds. Indeed, high schools are the environment where students interact with their peers, develop their capabilities, and form their expectations on their university careers. All these elements interact with established networks between high schools and *local* universities and the presence of promotion activities to inform students' choices.

Moving from this framework, our analysis uses the administrative data on Italian university students to infer the factors that drive students' mobility decisions toward *local* or *non-local* universities. This is done by comparing the choices of students that have attended the same high schools and applying a set of CCMNL that jointly considers the clustering of students in macro areas of residence, high schools, and the heterogeneity in their choices that depends on the chosen field of study. Moreover, we define students' mobility choices using a non-deterministic approach to account for the uncertainty in the definition of students' local area, overcoming issues related to the arbitrariness of administrative boundaries and thresholds.

Our results indicate that the macro area of residence is one of the most important determinants of students' choices concerning students' characteristics. The probability of choosing a *free non-local* alternative is higher for students from the South, the North-East or the Islands than for those living in central regions of the country and in the North-West. Moreover, we report evidence that the effect of students' characteristics is similar for *local* and *forced non-local* and, therefore, that these two alternatives attract the same kind of students. An exception is given by high school commuters that, on average, are more oriented toward *forced non-local* alternatives.

A different picture arises from the analysis of the high school effect, which provides innovative findings in understanding the role that schools play in influencing students' choices. Indeed, our results indicate that high schools have a different propensity to refer their students to *local* and *forced non-local* universities. While the differences between high schools explain between the 24.38% (North-West) and the 53.41% (Islands) of the unexplained variability in students' choices in the contrast between *local* and *free non-local* universities, this effect is lower when considering *forced non-local* universities. In this case, differences between high schools explain between the 9.17% (North-West) and the 35.55% (Islands) of the unexplained variability. This difference is common to all the Italian macro areas. Moreover, the high schools' influence on students' choices is stronger than the one associated with the chosen field of study only when considering the *local* alternative, with the exception of students in North-West regions, where differences in the field of study explain 41.55% of the unexplained variability. In general, the differences among fields account for a percentage of the unexplained variability that ranges from 4.17% in Islands to 41.55% in the North-West.

These results show that high schools are relevant in affecting students' choices, especially when considering *local* universities. Moreover, the weak correlation between the posterior estimates distributions of the random effects at the high school/curriculum level indicates that, from high schools' point of view, *local* and *forced non-local* universities are considered two different alternatives: schools that incentive students toward *local* alternatives do not also suggest *forced non-local* universities. This element can be interpreted as evidence of strong relationships between schools and nearby universities.

As for the differences between macro areas of residence, the results show that for students from southern regions and Islands, even a presence of a small positive effect toward *local* universities can counteract the general propensity of these students to prefer *free non-local* alternatives with respect to students in the other regions. Instead, students in central regions or the North-West prefer a *local* university even in the absence of a positive high school effect. For these students, *free non-local* and *forced*

non-local alternatives become relevant only when considering schools with a strong attitude toward these two categories. Finally, students residing in the North-East seem to prefer *free non-local* alternatives on average and that are less sensitive to the high school effect. Indeed, students in high schools located in these regions are indifferent between *local* and *free non-local* alternatives only in high schools that present a stronger than average high school effect toward *local* alternatives.

With respect to the distribution of the high school effects in the country, our results indicate that high schools have a higher propensity toward *local* alternatives when considering provinces with cities that host the most important universities in terms of students such as, for example, Milan, Turin, Naples and Rome. These elements suggest that high schools may contribute to polarizing students' choices in these areas of the country. However, policies that improve the relationship between high schools and *local* universities may help in retaining students in their local area.

Finally, the approach has shown that the results are similar across different specifications even when choosing different values for the thresholds δ_m used to define students' mobility choices. These results allow us to consider the main findings from this analysis robust to the uncertainty in the definition of students' local areas.

As far as the authors know, there are several elements of novelty in the analysis proposed. Namely, we advance a robust definition of students' local area, considering university accessibility and the supply of tertiary education institutions instead of administrative geographical boundaries. This approach should overcome the possible bias due to the absence of information on students' travel habits in the administrative archive by providing multiple classifications of students' choices according to their travel distance. Moreover, it is the first contribution that assesses the role of high schools' policies in influencing students' university choices toward *non-local* universities, which relies on Italian administrative data related to more academic years. The main limit of the analysis is related to the lack of information on students' socioeconomic profiles and, thus, the impossibility of disentangling the school effect from other confounding factors related to students' and families' income and wealth or to parents' profession and education. However, the MOBYSU.IT database contains also information on INVALSI's Economic, Social and Cultural status (ESCS) for all the students enrolled in an Italian university between 2018 and 2019. This indicator measures students' socio-economic and cultural background based on parents' education and employment, and the availability of specific resources at home. Since this variable is observed only in two cohorts, we are not able to use it to account for students' socio-economic backgrounds in the estimation. Nevertheless, to investigate the presence of a relationship between students' socio-economic background and the high school effect we have gathered the data regarding the average ESCS for all the high schools that are observed both in these two cohorts and in our database. The results show that the random effects that measure the high school effects are not correlated with high schools' average ESCS. Despite this result is not conclusive, it supports the robustness of the main finding indicating the high schools' effect on students' choice is not related to students' socio-economic background.[9]

---

[9] The results are available from the authors upon request.

Appendix A

See Tables 8, 9 and 10.

Table 8 ISCED-F 2013 fields of study

ISCED-F 2013 classification
Social sciences, journalism and information
Arts and humanities
Engineering, manufacturing and construction
Information and Communication Technologies
Business, administration and law
Natural sciences, mathematics and statistics
Services
Education
Health and welfare
Agriculture, forestry, fisheries and veterinary

The table reports the names of each ISCED-F 2013 fields of study

Table 9 High school curricula

High School Curricula in Italy
Lyceum
Artistic
Scientific
Linguistic
Humanities
Classical
Technical institutes
Naval
Surveyor
Aeronautical
Accounting
Industrial
Vocational institutes
Professional
Social
Agriculture

The table reports the names of each High School Curricula provided in Italy along with their classification in *liceo*, technical institutes and vocational institutes

Table 10 Distribution of students by region of residence

Variable	Proportion
North East	
Trentino Alto Adige	0.012
Veneto	0.083
Friuli Venezia Giulia	0.020
Emilia Romagna	0.071
North West	
Piedmont	0.069
Aosta Valley	0.002
Lombardy	0.159
Liguria	0.025
Centre	
Tuscany	0.060
Umbria	0.015
Marche	0.029
Lazio	0.105
South	
Abruzzo	0.024
Molise	0.006
Campania	0.111
Apulia	0.066
Basilicata	0.010
Calabria	0.031
Islands	
Sicily	0.080
Sardinia	0.024

The table reports the proportion of students in each region of residence

Author Contributions All authors contributed to the study conception and design. Material preparation, data collection, empirical analysis were performed by Cristian Usala. Isabella Sulis worked on the introduction and theoretical background. Mariano Porcu critically reviewed the manuscript. All authors read and approved the final manuscript and contributed to the conclusions.

directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Attanasio M (2022) Students' university mobility patterns in Europe: an introduction. Genus 78:17. https://doi.org/10.1186/s41118-022-00164-8

Attanasio M, Enea M (2019) La mobilitá degli studenti universitari nell'ultimo decennio in italia. In: Santis GD, Pirani E, Porcu M (eds) Rapporto sulla popolazione. L'istruzione in Italia, Il Mulino, Bologna, pp 43–58

Ballarino G, Colombo S, Panichella N, Piolatto M (2022) Human capital dynamics: the geographical mobility of high-school graduates towards university in Italy. Reg Stud 56(6):921–939. https://doi.org/10.1080/00343404.2021.1912723

Barone C, Assirelli G (2020) Gender segregation in higher education: an empirical test of seven explanations. High Educ 79:55–78. https://doi.org/10.1007/s10734-019-00396-2

Biancardi D, Bratti M (2019) The effect of introducing a Research Evaluation Exercise on student enrolment: evidence from Italy. Econ Educ Rev 69:73–93. https://doi.org/10.1016/j.econedurev.2019.01.001

Bratti M, Verzillo S (2019) The 'gravity' of quality: research quality and the attractiveness of universities in Italy. Reg Stud 53(10):1385–1396. https://doi.org/10.1080/00343404.2019.1566701

Castleman BL, Long BT (2016) Looking beyond enrollment: the causal effect of need-based grants on college access, persistence, and graduation. J Labor Econ 34(4):1023–1073. https://doi.org/10.1086/686643

Cattaneo M, Horta H, Malighetti P, Meoli M, Paleari S (2017a) Effects of the financial crisis on university choice by gender. High Educ 74(5):775–798. https://doi.org/10.1007/s10734-016-0076-y

Cattaneo M, Malighetti P, Meoli M, Paleari S (2017b) University spatial competition for students: the Italian case. Reg Stud 51(5):750–764. https://doi.org/10.1080/00343404.2015.1135240

Chetty R, Hendren N (2018) The impacts of neighborhoods on intergenerational mobility. Q J Econ 133(3):1107–1162. https://doi.org/10.1093/qje/qjy007.Advance

Chetty R, Friedman JN, Saez E, Turner N, Yagan D (2020) Income segregation and intergenerational mobility across colleges in the United States. Q J Econ 135(3):1567–1633. https://doi.org/10.1093/qje/qjaa005

Ciriaci D (2014) Does university quality influence the interregional mobility of students and graduates? The case of Italy. Reg Stud 48(10):1592–1608. https://doi.org/10.1080/00343404.2013.821569

Columbu S, Porcu M, 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-Econ Plan Sci 74(100):926. https://doi.org/10.1016/j.seps.2020.100926

Contini D, Tommaso MLD, Mendolia S (2017) The gender gap in mathematics achievement: evidence from Italian data. Econ Educ Rev 58:32–42. https://doi.org/10.1016/j.econedurev.2017.03.001

D'Agostino A, Ghellini G, Longobardi S (2019a) Exploring determinants and trend of stem students internal mobility some evidence from Italy. Electron J Appl Stat Anal (EJASA) 12(4):826–845. https://doi.org/10.1285/i20705948v12n4p826

D'Agostino A, Ghellini G, Longobardi S (2019b) Out-migration of university enrolment: the mobility behaviour of Italian students. Int J Manpower 40:56–72. https://doi.org/10.1108/IJM-07-2017-0169

Database MOBYSU.IT (Mobilità degli Studi Universitari in Italia) Research protocol MUR - Universities of Cagliari, Palermo, Siena, Torino, Sassari, Firenze, Cattolica and Napoli Federico II, Scientific Coordinator Massimo Attanasio (UNIPA), Data Source ANS-MUR/CINECA

Declercq K, Verboven F (2018) Enrollment and degree completion in higher education without admission standards. Econ Educ Rev 66:223–244. https://doi.org/10.1016/J.ECONEDUREV.2018.08.008

Dotti NF, Fratesi U, Lenzi C, Percoco M (2013) Local labour markets and the interregional mobility of Italian University Students. Spat Econ Anal 8(4):443–468. https://doi.org/10.1080/17421772.2013.833342

Dwenger N, Storck J, Wrohlich K (2012) Do tuition fees affect the mobility of university applicants? Evidence from a natural experiment. Econ Educ Rev 31(1):155–167. https://doi.org/10.1016/j.econedurev.2011.10.004

Enea M (2018) From South to North? Mobility of southern italian students at the transition from the first to the second level university degree. In: Perna C, Pratesi M, Ruiz-Gazen A (eds) Studies in theoretical and applied statistics. SIS 2016. Springer Proceedings in Mathematics and Statistics, vol 227. Springer, Cham, pp 239–249. https://doi.org/10.1007/978-3-319-73906-9_22

Engberg ME, Wolniak GC (2010) Examining the effects of high school contexts on postsecondary enrollment. Res High Educ 51(2):132–153. https://doi.org/10.1007/s11162-009-9150-y

Foot DK, Pervin B (1983) The determinants of postsecondary enrolment rates in Ontario. Can J High Educ 13(3):1–22. https://doi.org/10.47678/cjhe.v13i3.182904

Genova VG, Tumminello M, Enea M et al (2019) Student mobility in higher education: sicilian outflow network and chain migrations. Electron J App Stat Anal 12(4):774–800. https://doi.org/10.1285/I20705948V12N4P774

Giambona F, Porcu M, Sulis I (2017) Students mobility: assessing the determinants of attractiveness across competing territorial areas. Soc Indic Res 133(3):1105–1132. https://doi.org/10.1007/s11205-016-1407-1

Gibbons S, Vignoles A (2012) Geography, choice and participation in higher education in England. Reg Sci Urban Econ 42(1–2):98–113. https://doi.org/10.1016/J.REGSCIURBECO.2011.07.004

Hossler D, Gallagher KS (1987) Studying student college choice: a three-phase model and the implications for policymakers. Coll Univ 62:207–221

Hübner M (2012) Do tuition fees affect enrollment behavior? Evidence from a 'natural experiment' in Germany. Econ Educ Rev 31(6):949–960. https://doi.org/10.1016/j.econedurev.2012.06.006

Impicciatore R, Tosi F (2019) Student mobility in Italy: the increasing role of family background during the expansion of higher education supply. Res Soc Strat Mobil 62(100):409. https://doi.org/10.1016/j.rssm.2019.100409

Krezel J, Krezel ZA (2017) Social influence and student choice of a higher education institution. J Educ Cult Soc. https://doi.org/10.15503/jecs20172.116.130

Leckie G (2013) Module 11: three-level multilevel models - Stata Pratical. LEMMA VLE 11:1–47

Leckie G, Charlton C (2013) Runmlwin: a program to run the MLwiN multilevel modeling software from within Stata. J Stat Softw 52(11):1–40. https://doi.org/10.18637/jss.v052.i11

Long BT (2004) How have college decisions changed over time? An application of the conditional logistic choice model. J Econom 121(1–2):271–296. https://doi.org/10.1016/J.JECONOM.2003.10.004

Long JS, Mustillo SA (2021) Using predictions and marginal effects to compare groups in regression models for binary outcomes. Sociol Methods Res 50(3):1284–1320. https://doi.org/10.1177/0049124118799374

Oggenfuss C, Wolter SC (2019) Are they coming back? The mobility of university graduates in Switzerland. Rev Reg Res 39:189–208. https://doi.org/10.1007/s10037-019-00132-4

Pigini C, Staffolani S (2016) Beyond participation: do the cost and quality of higher education shape the enrollment composition? The case of Italy. High Educ 71(1):119–142. https://doi.org/10.1007/s10734-015-9892-8

Pitzalis M, Porcu M (2015) Passaggio a Nord. Come si ristruttura il campo universitario italiano? Scuola democratica 3:711–722. https://doi.org/10.12828/82088

Porcu M, Sulis I, Usala C (2021) CLADAG 2021 BOOK OF ABSTRACTS AND SHORT PAPERS. Proceedings e report, vol 128. Firenze University Press, Florence. https://doi.org/10.36253/978-88-5518-340-6

Rizzi L, Grassetti L, Attanasio M (2021) Moving from north to north: how are the students' university flows? Genus 77:8. https://doi.org/10.1186/s41118-021-00116-8

Rubin DB (1987) Multiple imputation for nonresponse in surveys, vol 81. Wiley, New York. https://doi.org/10.1002/9780470316696

Salter AJ, Martin BR (2001) The economic benefits of publicly funded basic research: a critical review. Res Policy 30(3):509–532. https://doi.org/10.1016/S0048-7333(00)00091-3

Santelli F, Scolorato C, Ragozini G (2019) On the determinants of student mobility in an interregional perspective: a focus on Campania region. Ital J Appl Stat 31(1):119–142. https://doi.org/10.26398/IJAS.0031-007

Santelli F, Ragozini G, Vitale MP (2022) Assessing the effects of local contexts on the mobility choices of university students in Campania Region in Italy. Genus 78:5. https://doi.org/10.1186/s41118-021-00144-4

Snijders T, Bosker R (2011) Multilevel analysis: an introduction to basic and advanced multilevel modeling, 2nd edn. Sage, London

Spiess CK, Wrohlich K (2010) Does distance determine who attends a university in Germany? Econ Educ Rev 29(3):470–479. https://doi.org/10.1016/J.ECONEDUREV.2009.10.009

Suhonen T (2014) Field-of-study choice in higher education: does distance matter? Spat Econ Anal 9(4):355–375. https://doi.org/10.1080/17421772.2014.961533

Türk U (2019) Socio-economic determinants of student mobility and inequality of access to higher education in Italy. Net Spat Econ 19(1):125–148. https://doi.org/10.1007/s11067-019-09445-w

UNESCO Institute for Statistics (2014) Isced fields of education and international standard classification of education 2011. Montrèal

Valero A, Van Reenen J (2019) The economic impact of universities: evidence from across the globe. Econ Educ Rev 68:53–67. https://doi.org/10.1016/J.ECONEDUREV.2018.09.001