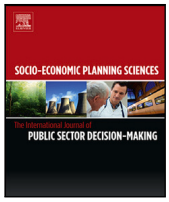
Socio-Economic Planning Sciences 94 (2024) 101926

Contents lists available at ScienceDirect

Socio-Economic Planning Sciences

journal homepage: www.elsevier.com/locate/seps



The more the better? How degree programs' variety affects university students' churn risk

Cristian Usala [a,*], Ilaria Primerano [b], Francesco Santelli [c], Giancarlo Ragozini [d]

[a] University of Cagliari, Department of Social and Political Sciences, Viale S. Ignazio, 78, Cagliari, 09123, Italy
[b] National Research Council of Italy, Institute for Research on Population and Social Policies, Corso San Vincenzo Ferreri, 12, Fisciano, 84084, Italy
[c] University of Trieste, Department of Economics, Business, Mathematics and Statistics, Via Alfonso Valerio, 4/1, Trieste, 34127, Italy
[d] University of Naples Federico II, Department of Political Sciences, Via Leopoldo Rodinò, 22, Naples, 80133, Italy

ARTICLE INFO

Keywords:
Students churn risk
Poisson Pseudo Maximum Likelihood
University policies

ABSTRACT

Students' intra- and, especially, inter-degree relocation poses significant challenges that tertiary education systems and universities must address to identify the best possible solutions for its reduction. Indeed, this phenomenon has a major impact on several aspects, from the reallocation of financial resources to factors related to the organization of spaces and activities, as well as in terms of human and social capital. Moreover, it is much more challenging when involving different universities in different geographical areas. This paper addresses the issue of churn risk among students and its relationship with the institutions' educational offerings, in terms of degree programs, by accounting for the impact of the characteristics of both universities and hosting areas. In particular, the analysis relies on Italian administrative data regarding students who decide to change university or field of study when enrolling in their second year of career. The results obtained performing the Poisson Pseudo Maximum Likelihood estimation show that the variety and the dimension of universities' services supply contribute to reducing students' churn risks and are positively associated with universities' retention rates.

1. Introduction

Risk analysis and management have taken on a crucial role in different types of enterprises, both public entities and private companies. These practices help identify adverse events and their associated losses, and develop a plan to mitigate them. While less developed, risk assessment practices exist at different levels, including higher education systems. In general, institutions, staff members, and students have limited awareness of the range of risks to which they and their environment are exposed [1].

Moreover, the importance of risk assessment practices has been highlighted by the wide circulation of new public management policy ideas and technologies within the globalized education space [2] that, over the last decade, has introduced business models in universities [3]. This 'new managerialism' [4] has profoundly changed also the Italian university system while fostering competition among the institutions that are incentivized to attract more and the best students [5] in a customer-input technology perspective [6] where students are considered at the same time inputs of educational services and consumers [7]. Therefore, as in other sectors where risks can directly affect the quality and continuity of business processes, in the education sector, they can directly influence the continuity of higher education and research activities [1]. In fact, the concepts of risk in education have been borrowed from business theories, expressly referring to the similarity between the characteristics of relationships in a business network and those between students and universities. Similarly, the definition of student churn risk is also directly derived from the business concept of customer churn [8].

Specifically, as customers may cease doing business with one company and move towards a competitor, students may dropout or change university during their study careers. Furthermore, as for an enterprise, customer churn behavior can be easily associated with economic costs; the same is true for universities. The loss of students for a university is not only an economic cost in terms of lost tuition paid. Indeed, it is also associated with social costs related to, for example, countries' tertiary education attainments and human capital. In addition, for each institution, high churn rates imply the worsening of the indicators on which the allocation of regular funding funds partly depends, as well as in practical-organizational activities and space use destinations.

In particular, relying on academic risk assessments from students' perspectives means accounting for their decision to change their paths

* Corresponding author.
E-mail addresses: cristian.usala@unica.it (C. Usala), ilaria.primerano@irpps.cnr.it (I. Primerano), fsantelli@units.it (F. Santelli), giragoz@unina.it (G. Ragozini).

https://doi.org/10.1016/j.seps.2024.101926
Received 28 February 2023; Received in revised form 27 February 2024; Accepted 9 May 2024
Available online 11 May 2024

during their careers. This can occur when students, for a variety of reasons, decide to: (i) enroll in another university, (ii) change field of study, and (iii) drop out. So, all university governance have to tackle student churn risk to identify the elements related to higher risks of students' career disruption and those associated with better performances in terms of student retention. For this reason, data-driven decision-making has become a very important tool for university administrators who aim to increase student retention by improving their policies and structures and seeking to innovate the educational system and learning methodologies to enhance students' academic performance.

This paper deals with Italian students' decision to churn during their three- and five-year degree programs, focusing on changes occurring between the first and the second year of their university career. Specifically, based on students' relocation among universities and fields of study, we analyze two main features of institutions' churn risk. The first depends on the number of students that change university and is related to a direct economic cost for the universities. The second is related to the number of students that change their field of study while remaining in the same university, which is connected to organizational and social costs. Our aim is to understand the role that universities' characteristics have in reducing the loss of students related to churn decisions by focusing on the role played by the dimension and the variety of their offer of degree programs.

To account for the relevant share of cases where we do not observe students who decide to churn (i.e., zero values in the dependent variables), the analysis is based on the Poisson Pseudo Maximum Likelihood (PPML) estimation technique [9]. The results indicate that the number of degree programs offered by universities, intended as a measure of the dimension and the variety of opportunities provided by institutions, contributes to reducing churn rates, especially when considering students who change university during their first year of career. Also, a higher number of professors and administrative staff per student increases the retention rate, indicating a positive relationship with the quality of teaching and student services provided by universities.

This paper is structured as follows. Section 2 presents an overview of the Italian university institutional setting (Section 2.1) together with the theoretical background underpinning this research (Section 2.2); Section 3 provides an overview of the data; Section 4 presents the methodological approach; Section 5 shows and discusses the results obtained by the implementation of the PPML estimator; Section 6 provides some conclusion remarks and discusses the policy implications of our results.

2. Background

2.1. Italian higher education framework

The Italian university system is embedded in a heterogeneous framework consisting of countries of the European Union [10,11]. In this context, several different policies have been adopted by countries, according to their own political aims and financial availability in expenditures for higher education [12]. In particular, following the EU convergence process, the Italian higher education system has profoundly changed in the last decades. In particular, two elements, among others, have concurred to foster the competition for students among institutions. In 1993, the law 537 on the financial autonomy of universities was introduced to reduce universities' expenditures by introducing strong elements of competitiveness between departments and universities. Then, in 1999, the "Bologna Process" proposed the realization of the European Higher Education Area (EHEA), which intended to increase cooperation, knowledge, and mobility for researchers and students in Europe. These elements have concurred in raising the internal competition for students among universities. Indeed, in the last years, several authors explicitly pointed out how the whole framework has shifted towards a highly competitive approach [13,14]. Embedding itself in what is called *new public management* [15], universities act as actors in a quasi-market fashion [16] where they compete to gain public funding assigned to higher education.

Therefore, understanding the factors that drive students' decisions about their university careers, including their choice of degree programs and institutions, is critical for universities. In particular, student retention is one of the most important challenges to which Italian educational institutions are exposed. Indeed, this phenomenon covers such a prominent role that it has also been considered by the Italian National Agency for the Evaluation of Universities and Research Institutes (ANVUR) among the set of indicators of the Italian Higher Education Quality Assurance system called AVA (Autovalutazione, Valutazione periodica, Accreditamento, i.e., Self-assessment, Periodic Evaluation, Accreditation) that operates annually since 2013.[1] ANVUR uses a formula-based element in order to achieve an assessment of university performances and, in turn, assign resources accordingly [17].

In the AVA system, the regularity of university studies is one of the five dimensions related to students' careers which universities include in their annual monitoring report. The specific set of indicators about the course of study and regularity of careers is defined at both degree programs (Annex 6.1 of AVA guidelines) and university levels (Annex 5 of AVA guidelines). It includes the same four indicators: (i) percentage of students who continue their careers in the university system in the 2nd year; (ii) percentage of matriculated students at a bachelor's, master's or single-cycle degree program who graduate, in the same degree program, within the standard duration of the course; (iii) percentage of freshmen at a bachelor's, master's, or single-cycle degree program who continue their careers in the second year in a different degree program of the university system; and (iv) percentage of degree program dropouts after $N + 1$ years. In addition to these, there is a fifth indicator observed only at the university level, it measures the percentage of students who continue their careers in the university system in the 2nd year at the same university. A specific set of indicators designed to support the self-evaluation and review activities of courses delivered by telematic universities is also added to the indicators already available.[2] Furthermore, indicators of student retention and, therefore, churn risk, are also found among the specific objectives of the National Plan for Science Degrees (PLS — Piano Lauree Scientifiche) and the Tutoring and Orientation programs (POT — Piani di Orientamento e Tutorato) implemented by Italian universities to improve the completion of study courses through targeted actions that guide high school students to consciously choose the most suitable university degree program according to their abilities.

In this framework, reducing the phenomenon of university dropout and churn risk is one of the main policy goals of universities governance that are interested in reducing both aspects and encouraging students to undertake a regular academic path.

2.2. Literature review

Understanding students' educational decisions is a strategic factor for universities and policymakers at both university and local government levels given the impact that these choices have on future migration patterns of highly educated individuals [18,19], regional competitiveness and growth capacity [20,21], and the allocation of public funds that depends on universities' effectiveness in attracting students from other areas of the country [22].

[1] The AVA has been introduced by the Ministerial Decree 1154/2021 "Self-assessment, assessment, initial and periodic accreditation of venues and courses of study" - Decreto Ministeriale del 14 ottobre 2021, n. 1154 "Autovalutazione, valutazione, accreditamento iniziale e periodico delle sedi e dei corsi di studio".

[2] For details see https://www.anvur.it/attivita/ava/indicatori-di-monitoraggio-autovalutazione-e-valutazione-periodica/indicatori-cds-e-atenei-telematici/.

Despite the importance of the phenomenon and the set of measures implemented by universities to address students' churn risk and to improve their retention, recent researches have identified important risk factors in dropout [23,24] and churn decisions [25,26]. Furthermore, other studies have focused on students' mobility choices and university attractiveness by analyzing students' migration flows within Italy [27–29] and abroad [30], with specific reference to students participation in Erasmus programs [31,32]. With respect to the Italian context, several studies have analyzed the push and pull factors for university students' mobility [33,34] mainly tracing a student migration flow from southern to northern regions of the country that mirrors the labor migration one [35–39]. These studies have mainly focused on students mobility decisions in the transition from high school to university [40], highlighting how even during high school a few hints about the future choices of students are available due to proficiency tests [41,42], and from bachelor's to master's degree programs [43,44], while only a few studies have analyzed the transition from one year to the next in the same degree program and/or university [45–47]. Moreover, students' educational choices and performances have been analyzed to identify universities and hosting territories that are good importers and/or good exporters [38,43,48], to analyze university student's careers by also looking at churn decisions [25,26,49], and to investigate the factors beside university dropout such as students' socio-demographic background, academic competencies, motivation, social and academic integration [23,50]. With respect to this last aspect, the literature highlights an important role in students' educational choices of several elements related to students' age, sex, past mobility choices, socio-economic and educational background, and family characteristics [51–53]. In addition, recent studies have shown that students' educational choices are also affected by universities' characteristics, such as research quality [54], financial aid [55], and hosting areas' characteristics [56].

Aina et al. [52], by comparing all the countries covered by OECD statistics with respect to the non-completion and first-year dropout rates of students enrolled in bachelor degree programs, have also highlighted the widespread of the university withdrawal phenomenon across countries and how important this phenomenon is even abroad. This problem, widely perceived even worldwide, has prompted researchers to develop theories and methods in an attempt to understand [57] and to predict student churn [58] as well as to identify at-risk student and the potential financial, academic, and/or personal factors leading them to dropout [59] by means of data mining tools and statistical techniques.

Moving from this framework, this paper investigates the relationship between universities' and hosting areas' characteristics with undergraduate students' attitudes to change the field of study and/or university in the transition from the first to the second year of their career. Inter-degree students relocation or churn decisions can be seen as a risk for sending institutions and an opportunity for receiving ones. Indeed, from a university perspective, churn rates can be used to measure their ability to meet students' needs and expectations. At the same time, it is a risk factor that should not be underestimated for sending institutions since it is related to a net loss of students. This risk is even more pronounced when students decide to change universities and not only disciplinary fields.

Instead, from students' perspective, churning is a peculiar kind of risk in their university careers. Indeed, by changing university or degree programs after one year, students who are struggling in their current field of study or university may correct their careers in time to solve their academic issues and improve their performances and educational attainments.

3. Data

To analyze the churn phenomenon and its different aspects, we propose a data integration process [60] that is a challenging but exhaustive way to obtain information about a complex and multi-level problem. In order to do so, we combined different sources of secondary data regarding students' university careers, universities' features, and hosting areas' labor market and socio-economic characteristics. Specifically, the data regarding students' university careers is extracted from the administrative database MOBYSU.IT[3] which includes the administrative data on the population of students enrolled in Italian universities [61]. The information on universities' and hosting areas' characteristics is obtained from USTAT[4] and ISTAT databases.

In particular, from MOBYSU.IT we have extracted the information regarding the population of students enrolled in a three- or five-years degree program in an Italian university between the academic years 2010/2011 and 2018/2019. In order to follow students for a period of time in which the churn decision can reasonably happen, we collected the information regarding students' first two years of career with respect to the chosen university and the degree program. Following this strategy, we observe 2,345,590 students enrolled in 67 different degree programs that are grouped in 14 CUN areas following the classification proposed by the Italian National University Council (CUN), which groups the degree programs in disciplinary areas or fields of study according to the similarity of their contents.[5][6] We decided to use this classification since, besides being institutionalized by law in the Italian Education System, it also has two main advantages compared with other classifications (e.g., the 'International Standard and Classification of Education: Field of Education and Training' from UNESCO). First, it is used in USTAT to define the number of professors and researchers for each disciplinary area. Therefore, it allows to have the same level of detail in the classification of degree programs between the two databases. Second, CUN areas follow a similar pattern to the partition within universities *Academic Departments*. Usually, each department is devoted to a particular academic discipline, so CUN areas and departments tend to overlap to a large extent. Furthermore, each department also has some academic, political, and financial autonomy. Thus, there is also a sort of internal competition among departments belonging to the same university. Indeed, from a department's perspective, if students decide to leave one of the degree programs they are providing, it is more beneficial to retain them in a different course within the same department.

Since we are interested in the relationship between students' inter-degree relocations between the first and the second year of their career and universities' characteristics, we define our population of interest by following three steps. First, we do not consider data on 282,820 students not enrolled in any university during their second year (e.g. dropouts). Indeed, these students have left the Italian university system and we cannot trace their destination. Second, we do not include the data on 4129 students enrolled in universities for which USTAT does not provide information on their characteristics.[7] Third, we do not consider 2667 students enrolled in the Defence Studies degree

[3] Data drawn from the Italian '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' carried out by the University of Palermo (head of the research program), the Italian 'Ministero Università e Ricerca', and INVALSI.

[4] See http://ustat.miur.it/.

[5] See https://www.cun.it/cun/comitati-d-area/.

[6] The degree programs are defined according to Italian '*Classi di Laurea*' classification. *Classi di Laurea* group the specific courses provided in the Italian university system in broader categories based on their disciplinary contents. For example, the *Classe di Laurea* 'L-18' contains the courses related to business studies that can be more oriented towards marketing studies (course in business and marketing) or, for example, management studies (course in business and management). This classification is also used in public tenders to distinguish between different profiles of graduated students.

[7] These universities are the San Raffaele of Rome, the Link Campus University, the Humanitas University, and the Rome Saint Camillus University.

Table 1
Descriptive statistics on students' churn decisions.

	Enrolled students		Churn							
	Total		Churn		CUN area		University		Univ. & CUN	
	N	%	N	%	N	%	N	%	N	%
Total	2,055,974		196,708	9.6	86,165	43.8	38,540	19.6	72,003	36.6
University macro area:										
North-West	544,342	26.5	44,640	8.2	18,330	41.1	8365	18.7	17,945	40.2
North-East	409,298	19.9	34,713	8.5	13,516	38.9	7626	22.0	13,571	39.1
Center	489,886	23.8	53,754	11.0	24,592	45.7	11,165	20.8	17,997	33.5
South	442,803	21.5	45,093	10.2	18,436	40.9	8368	18.6	18,289	40.6
Islands	169,645	8.3	18,508	10.9	11,291	61.0	3016	16.3	4201	22.7
CUN disciplinary area:										
Economic and statistical sciences	304,837	14.8	18,700	6.1	6501	34.8	6074	32.5	6125	32.8
Literary and art-historical sciences	247,107	12.0	15,239	6.2	5842	38.3	4220	27.7	5177	34.0
Industrial and information engineering	245,326	11.9	23,296	9.5	9555	41.0	4201	18.0	9540	41.0
Biology	197,493	9.6	48,531	24.6	23,754	48.9	4243	8.7	20,534	42.3
History, education and psychology	185,373	9.0	11,496	6.2	4916	42.8	2909	25.3	3671	31.9
Medical sciences	181,195	8.8	9769	5.4	1748	17.9	5747	58.8	2274	23.3
Law	176,705	8.6	16,825	9.5	6908	41.1	4575	27.2	5342	31.8
Political and social sciences	172,655	8.4	11,709	6.8	5238	44.7	1941	16.6	4530	38.7
Civil engineering and architecture	118,550	5.8	9326	7.9	3779	40.5	2246	24.1	3301	35.4
Agricultural and veterinary sciences	82,346	4.0	6749	8.2	2994	44.4	1095	16.2	2660	39.4
Mathematics and computer science	55,919	2.7	6110	10.9	3250	53.2	677	11.1	2183	35.7
Chemistry	38,319	1.9	9317	24.3	5717	61.4	260	2.8	3340	35.8
Geology	27,065	1.3	6356	23.5	3990	62.8	158	2.5	2208	34.7
Physics	23084	1.1	3285	14.2	1973	60.1	194	5.9	1118	34.0

Col 2 shows the total number of students between 2010 and 2018. Col 3 reports the share of students by university macro area and CUN disciplinary area. Col 4 and 5 shows the number of students' that decide to churn and the proportion of these students with respect to the total number of students. Col 6 and 7 show the number and the percentage of students who changed CUN area within the same university with respect to the number of churners in the group. Col 8 and 9 show the number and the percentage of students who decided to change university within the same CUN area. Col 10 and 11 show the number and the percentage of students who changed university and CUN area with respect to the number of churners in the group.

program because we cannot classify these courses in any CUN area. Following this strategy, we include the data on 2,055,974 students.

To classify students' churn decisions, we exploit the information on the chosen university and the CUN area. More specifically, we account for differences in universities' and hosting areas' characteristics by defining our unit of analysis as the combination of the university, the CUN area, and the province hosting the university. Following this strategy, we identify 1007 units. In each unit, we classify students' choices into three categories: stay, churn-in, and churn-out. Students are defined as stayers if they do not change the university or the CUN area between their first and second year of career. Instead, a churn-in decision is observed when students change the CUN area while staying in the same university between her/his first and second year of career, and a churn-out decision is observed when a student changes the university. Moreover, the students who decide to churn-out can be further divided into two groups depending on whether they have changed their CUN area or not. Therefore, the churn-out phenomenon is investigated considering two levels. In the first, the churn-out phenomenon is analyzed in general to shed light on how universities can reduce the losses of students due to their relocations. In the second, the two groups are separately analyzed to understand whether the determinants of churn-out flows change depending on whether the students have also decided to change their disciplinary area.[8]

To have more information on the dimension of the churn phenomenon in Italy, Table 1 provides descriptive statistics on students' churn decisions by university macro area and CUN disciplinary area.

From data depicted in Table 1, it is clear that students behave differently according to universities' geographical macro-area and CUN disciplinary area. The percentage of churning students ranges from 8.2% (North-West) to 11% (Center). The three CUN areas in which there is a dramatically high churn rate are Biology (24.6%), Geology (23.5%), and Chemistry (24.3%). Such trajectories are historically related to passages of students towards the CUN area of Medical Science, which interestingly shows also the lowest churn rate (5.4%). Medicine is not the only degree program with a mandatory admission test and a fixed number of students each year, but it is undoubtedly the most coveted. Many exams of the first year of Biology and Chemistry programs can be easily transferred into the Medicine program; thus, many students who cannot pass the Medicine admission test successfully enroll in Biology/Chemistry in the first year and try the Medicine test again one year later. A more complex churn behavior to interpret is Geology's high value. However, this area covers a very low share of enrolled students (1.3% of the total).

Table 1 also provides information on the type of observed churning behavior. Considering the percentage of students that change only the CUN area, we can notice a peculiar behavior in the Islands macro-area, where most students change only the CUN area after one year (61%, while for other macro-areas such share is around 39%) but are less encouraged to change also the university. This element can be related to higher migration costs for students in the Islands, which could reduce the number of affordable options available to them.

Overall, Medical Sciences (17.9%), Economic and statistical sciences (34.8%), and Literary and art-historical sciences (38.3%) show the lowest share of students changing CUN areas and the highest percentages of students that decide to change only the university while staying in the same area. This element suggests that these students prefer to stay in their area even when unsatisfied with their chosen university. However, for them, the churn phenomenon is more related to students who decide to change both universities and CUN areas. A different picture emerges when considering students in Geology, Chemistry, Physics, and Biology that show the lowest percentage of students that decide to change only the university while higher shares of students that decide to change the disciplinary area both within their university than by enrolling in a different institutions.

The MOBYSU.IT database also provides the information on our key variable of interest: the number of degree programs in the unit. This number is computed by counting the degree programs that are provided by the university in the CUN area in the province considered. Our aim is to estimate the effect of an increase in the variety of degree programs on universities' performance in terms of student retention. In other

[8] We thank the anonymous reviewer for this suggestion.

Table 2
Descriptive statistics on covariates.

Variables		Descriptive	
Name	Description	Average	Std dev.
University:			
Non academics/N	Number of non-academic employees/Total number of students	0.209	0.121
Telematic	Telematic (on-line) universities	0.043	0.203
Scholars/N	Number of professors and researchers/Total number of students	0.218	0.139
Revenues/N	University tax revenues/Total number of students	8354	9375
University/CUN area:			
N of degree programs	Number of *Classi di Laurea* in the unit	2.171	1.536
Professors in the area/N	Number of professors in the CUN area/Total number of students in CUN area	0.134	0.353
Hosting area:			
Unemployment	Provincial unemployment	0.108	0.053
Taxable income	Municipal total taxable income/number of tax-payers	19,574	3317
Regional GDP	Regional added value per capita	24,689	6264

The table reports the name and the descriptions of variables used in estimation along with the average and the standard deviation.

words, we aim to understand whether increasing the number and the variety of degree programs may help reduce students' churn risk.

The data regarding universities' characteristics is obtained from USTAT. USTAT is an open-data portal managed by the Italian Ministry of University and Research that contains information on the characteristics of Italian universities. In particular, for each university, we observe the number of non academic employees and scholars (professors and researchers), the amount of revenues per student that is gathered by the university, and the information on whether the institution is a telematic (online) university. At the unit level, we observe the information on the number of professors in the specific CUN area. This information is used to account for the different characteristics of universities' supply that may affect students' churn decisions besides the number of degree programs.

Finally, to account also for territorial differences across the country in terms of labor market opportunities and socio-economic conditions, we collected the data on the provincial unemployment rate and the regional value added per-capita from ISTAT, and the information on the provincial total taxable income per tax-payer from the web site of the Italian Ministry of Economy and Finance.[9] To account for potential simultaneity biases, all the covariates are lagged by one year. Moreover, since telematic universities do not have a physical location, we have imputed the average values of the hosting areas' characteristics of other universities.

Table 2 shows the definition and the descriptive statistics of all variables retrieved from the databases that are included in the estimation.

4. Methodological approach

To understand the effect of universities' characteristics on students' churn flows and inter-degree relocations, we define our outcome variables based on the definition of students' churn decisions. In principle, this analysis should be based on the micro-data on students' choices and characteristics. However, we chose to use the data at the aggregate level since our primary interest concerns the aggregate impact of degree programs' variety at the institutional level rather than its heterogeneous effects on different types of students. Therefore, using aggregate data enables a more straightforward interpretation of the results. Indeed, despite providing information on the heterogeneity of the effect of interest, individual data would require the estimation of non-linear discrete choice models (e.g., multinomial logit) that would complicate the interpretation of the aggregate impact at the institutional level. Therefore, we believe that using aggregate data is a reasonable and appropriate choice for our study, as it allows us to answer our research question in a clear and rigorous way. However, using micro-data could help to improve our understanding of the heterogeneity of the effect based on individual characteristics. As pointed out in Section 6, we leave this possibility for future research, limiting our analysis to the estimation of the effect at the aggregate level.

Based on these considerations, we define two outcome variables. The first outcome depends on the number of students that decide to change university between their first and second years (i.e., churn-out decision) and is defined as:

$$I_{out,i,t} = \frac{N_{out,i,t} \times 1000}{N_{stay,i,t}} \tag{1}$$

where $N_{out,i,t}$ is the number of students enrolled in unit i at time $t-1$ that decide to change university at time t and $N_{stay,i,t}$ is the number of students that is observed in the same unit i between year $t-1$ and year t. Therefore, $I_{out,i,t}$ is the number of students that decide to change the university between $t-1$ and t every 1000 students that stay in the unit between the two years.

The second outcome is defined based on the number of students that decide to change CUN area between their first and second year of career by staying in the same university (i.e., churn-in decision). This variable is defined as:

$$I_{in,i,t} = \frac{N_{in,i,t} \times 1000}{N_{stay,i,t}} \tag{2}$$

where $N_{in,i,t}$ is the number of students enrolled in the unit i in time $t-1$ that decide to change CUN area in time t. Hence, $I_{in,i,t}$ measures the number of students that, between year $t-1$ and year t, decide to change the CUN area for every 1000 students who stay in the unit in the same time frame.

Therefore, these two outcomes help to measure two different phenomena. In the first case, we measure universities' ability to retain their students between the first and second year in a context of competition between universities. In the second case, we measure units' ability to retain students in a context in which different CUN areas in the same university compete for students who decide to churn. Therefore, while in the first case, we have a net loss for the whole university, in the second case, we measure the loss of students for the unit considered.

Moreover, although these two phenomena may be measured by the number of students in each churn category, we opted to use the ratio between churners and stayers to account for the overall dimension of universities in terms of students. Indeed, universities with a higher number of enrolled students are more prone to observe a higher number of churners due to the fact that, for a given churn probability, bigger

[9] See https://www1.finanze.gov.it/finanze/pagina_dichiarazioni/public/dichiarazioni.php.

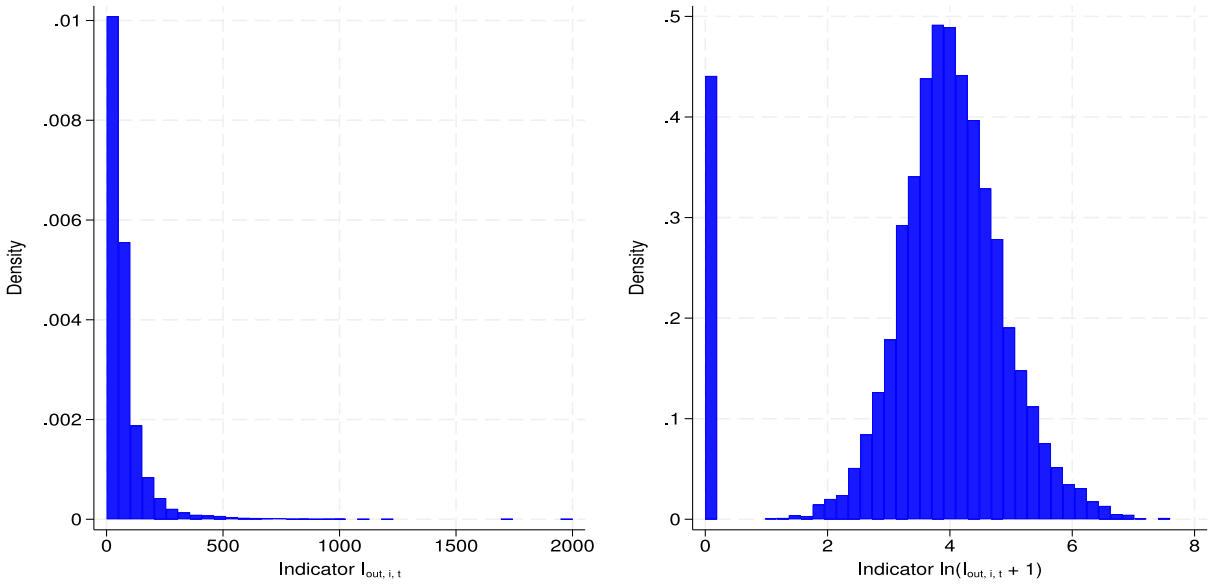


Fig. 1. Distribution of the indicator $I_{out,i,t}$.
The figure depicts the distribution of the indicator $I_{out,i,t}$ defined in Eq. (1). The first plot shows the distribution of the indicator in levels while the second depicts the distribution of the transformed indicator $\ln(I_{out,i,t}+1)$.

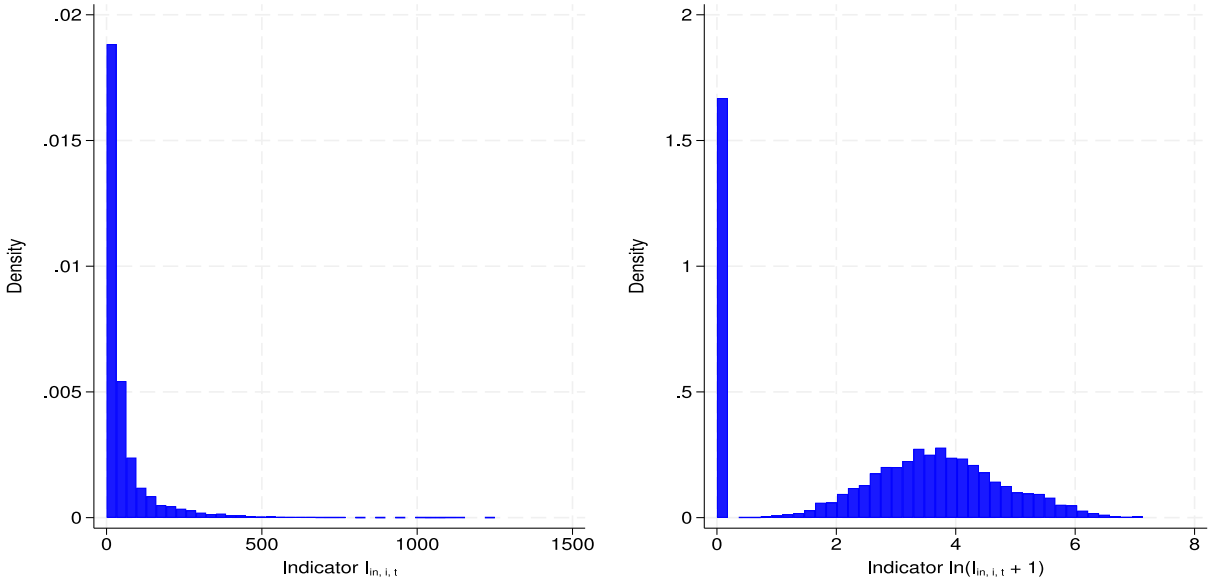


Fig. 2. Distribution of the indicator $I_{in,i,t}$.
The figure depicts the distribution of the indicator $I_{in,i,t}$ defined in Eq. (1). The first plot shows the distribution of the indicator in levels while the second depicts the distribution of the transformed indicator $\ln(I_{in,i,t}+1)$.

universities will have a higher number of churners. Therefore, by using a regression of the number of churners on units' characteristics we will not be able to disentangle the effect that these elements have on the number of enrolled from the one that they have on the number of churners. Instead, by using a ratio, we have a relative indicator that helps measure the importance of the churn phenomenon by accounting for the dimension of the unit.

Figs. 1 and 2 show the distributions of the outcome variables considering two different formulations. The first (on the left) considers the dependent variables in levels without any transformation. The second (on the right) shows how the distribution changes when we consider, respectively, $\ln(I_{out,i,t}+1)$ and $\ln(I_{in,i,t}+1)$. This transformation is used to show the share of zeros in the data and the distribution of the data when we consider logs.

As shown in the two Figures, one important characteristic of the outcome variables defined in Eqs. (1) and (2) is the relevant number of zeros observed in the data.[10] Indeed, 701 observations, equal to the 8.5% of the total, in our data have a value of $I_{out,i,t}$ equal to zero. This share is even higher when considering the index $I_{in,i,t}$ for which we observe 2489 (30.5%) zeros.

To account for this element, the outcome variables are modeled using the Poisson Pseudo Maximum Likelihood estimator (PPML) [9]. Indeed, the use of the PPML entails various advantages over other alternative techniques, such as the estimation of log-linearized equations with ordinary least squares (i.e., the transformed data in Figs. 1 and 2). First, given that the outcome variables are specified in levels and not

[10] We observe a zero when the number of churners in the unit is zero.

in logs, this technique allows to have zeros in the dependent variables. Indeed, as shown in Santos Silva and Tenreyro [62], the PPML performs better than other approaches even when the dependent variable contains a high number of zeros. Moreover, as noted, for example, in the health econometrics literature (see Manning and Mullahy [63], Powell and Seabury [64]), using a log-linearized specification may lead to wrong interpretations of the results due to Jensen's inequality. Since $\ln(E[y|X]) \neq E[\ln(y)|X]$, a log-linear specification may inform on the role that X has on the logs of the ratio between churners and stayers rather than on the ratio itself. Another important advantage is that Poisson regressions do not require additional assumptions with respect to the OLS. As highlighted by Santos Silva and Tenreyro [9], the only assumption required by the PPML is that the conditional expectation of the dependent variable in Eq. (3) is correctly specified. Under this condition, the PPML represents a robust estimation technique with respect to the OLS associated with the log-linear specification that, in the presence of heteroskedasticity, is not biased only under very specific conditions (see Santos Silva and Tenreyro [9] in page 644). Moreover, the PPML requires fewer assumptions than other techniques that are used to model dependent variables with zeros such as the negative binomial that requires, besides defining the conditional expectation of the dependent variable, to specify also the variance.[11] Indeed, as shown in Gourieroux et al. [66], Poisson regressions do not require any distributional assumption for the dependent variable and can be used to model any dependent variable with nonnegative values (see also Correia et al. [67]). This feature also rules out other possible solutions such as two-part models or hurdle models (e.g., Cragg [68], Mullahy [69]) that require additional assumption on the distribution of the data [70].[12][13]

Within the PPML framework, we model the ratios between the number of churners and stayers as follows:

$$I_{j,i,t} = \exp(X_{i,t}\beta + \gamma_i + \omega_i + \delta_i) + \epsilon_{j,i,t} \quad \text{with} \quad j = out, in \tag{3}$$

where j indicates which dependent variable is considered in the estimation, i indicates the unit given by the combination of university, hosting province, and CUN area, and t the academic year. Moreover, $X_{i,t}$ is a set of universities' and hosting areas' characteristics that affect the number of churners in the unit, γ_i is the set of university fixed effects that account for all the time-invariant elements that affect our outcomes and vary among universities, ω_i is the set CUN area fixed effects that account for all the characteristics that are specific to the considered CUN area and do not vary over time, and δ_i is the set of regional fixed effects that account for all time-invariant characteristics of the hosting regions. To allow the presence of universities and CUN areas fixed effects, we use the STATA routine PPMLHDFE developed by Correia et al. [67] to account for the presence of high-dimensional fixed effects. As will be discussed in Section 5, we will estimate various specifications using different combinations of fixed effects. Therefore, depending on the considered specification, we will consider two sources of variation: the variation within universities and the one within CUN areas. The use of fixed-effects estimation over a random-effect multilevel estimator is grounded in the ability of fixed-effects to account for the presence of time-invariant unobserved heterogeneity that could introduce a correlation between our regressors and the error term, potentially leading to biased results [65]. Moreover, as we are not primarily concerned with modeling the variability of churn flows, we believe that fixed effects offer a superior approach to estimating the average effect of our variable of interest while controlling for all time-invariant factors that may influence our analysis.

The set of variables $X_{i,t}$ contains information on universities and hosting areas. Our variable of interest is given by the number of degree programs provided in the unit in year t. The associated coefficients measure how the variety in the supply of degree programs in the unit may help to reduce the risk of losing students due to churn decisions. However, since students' decisions may depend also on other elements besides the number of degree programs, we account for several other characteristics.

Specifically, at the macro level, for each university, we observe the number of non academic employees, the number of scholars, and the revenues obtained. The first characteristic measures the quality of non academic services provided to students and professors. The number of scholars is a proxy for the quality of the academic services provided by the universities. Indeed, given the number of enrolled students, a higher number of scholars means smaller classrooms, and it can be positively related to the time that each professor dedicates to each student. Finally, revenues are a measure of costs that students have to bear to enroll in the considered university. To account for the dimension of universities, each regressor is divided by the number of enrolled students.

At the unit level, for each CUN area, we observe the number of degree programs and the number of professors. While the first regressor is the variable of interest, the second helps to account for the variability in the number of professors among CUN areas in the same university.

Finally, at the hosting area level, we account for the provincial unemployment, regional added value, and taxable income. These elements help to account for the heterogeneity in labor market opportunities and socio-economic characteristics among different areas of the country. The descriptive statistics regarding the variables used in the estimation, along with their definition, are reported in Table 2.

Given the non-linear form of the model, all parameters are to be interpreted as semi-elasticities. As shown in Santos Silva and Tenreyro [9], for a given parameter β the semi-elasticity is obtained by computing $(\exp(\beta) - 1) \times 100\%$. For example, for a coefficient equal to 0.5, the expected increase in the conditional expectation of the dependent variable associated with a unit increase in the considered regressor is equal to $(\exp(0.5) - 1) \times 100\% = 64.87\%$.

5. Results and discussion

The PPML estimation results are shown in Tables 3 and 5. Each table holds the estimated coefficients, their standard errors, and the associated p-values.[14] In the following, only the coefficients for the variables attaining the university quality and organization will be analyzed, neglecting the variables that control the contextual effects related to hosting areas' characteristics and fixed effects.

Looking at the results for models in Table 3, we notice that the number of degree programs negatively affects the churn-out indicator (the associated β ranges from $-.115$ to $-.140$). This implies that increasing the variety of degree programs by one additional program reduces the churn indicator by at least 11%, increasing the retention rate. This result is also stable when we introduce the regional, university, and CUN area fixed effects. Also, the number of professors for disciplinary areas and administrative staff members have negative effects on churn, especially in models 5 and 6 that consider regional and CUN areas fixed effects: increasing the number of professors per student reduces the

[11] See, for example, Wooldridge [65] in Chapter 19.

[12] Another reason to avoid two-parts models in our case is given by the fact that we do not have any reason to suspect the existence of two separate processes for the extensive margin (i.e., whether the outcome variable is greater than zero) and the intensive margins (i.e., the number of students that decide to churn).

[13] PPML is widely employed in many contexts with nonnegative dependent variables that present a relevant mass of zeros such as, among the most notable, trade [9], environmental agreement [71], arms trade [72], crime and violence [73], and currency unions [74]. See Santos Silva and Tenreyro [70] for a review of the state of the art of the PPML applications and advantages.

[14] Since we have population data, the estimated coefficients should be interpreted as population parameters. However, standard errors can be used to assess the preciseness of our estimates.

Table 3
PPML estimation results for the $I_{out,i,t}$ indicator.

	(1)	(2)	(3)	(4)	(5)	(6)
N of degree programs	−0.116***	−0.121***	−0.123***	−0.115***	−0.140***	−0.116***
	(0.00830)	(0.00891)	(0.00871)	(0.00889)	(0.0105)	(0.0111)
Revenues/N (10k€)		−0.0361	−0.0919***	0.198**	−0.0316	0.234***
		(0.0259)	(0.0264)	(0.0620)	(0.0242)	(0.0594)
Scholars/N		1.525***	2.483***	−0.317	2.345***	−0.213
		(0.290)	(0.317)	(0.707)	(0.289)	(0.680)
Professors in the area/N		−0.0239	−0.0246	0.0174	−0.413**	−0.384**
		(0.105)	(0.103)	(0.0980)	(0.157)	(0.136)
Non academics/N		−1.093***	−0.842***	−1.419*	−1.026***	−1.510**
		(0.211)	(0.185)	(0.617)	(0.167)	(0.562)
Provincial Unemployment		−0.968*	−1.519***	−2.206***	−1.742***	−2.316***
		(0.420)	(0.459)	(0.528)	(0.414)	(0.466)
Taxable income per capita (10k€)		−0.128	−0.0618	−0.393***	−0.168*	−0.665***
		(0.0690)	(0.0749)	(0.0907)	(0.0746)	(0.0973)
Regional GDP per capita (10k€)		−0.0277	−0.269*	−0.0358	−0.326**	0.0346
		(0.0441)	(0.107)	(0.193)	(0.106)	(0.175)
Constant	4.549***	4.907***	5.236***	5.908***	5.812***	6.391***
	(0.0263)	(0.148)	(0.252)	(0.513)	(0.237)	(0.459)
CUN area FE	No	No	No	No	Yes	Yes
University FE	No	No	No	Yes	No	Yes
Region FE	No	No	Yes	Yes	Yes	Yes
Observations	8160	8160	8160	8160	8160	8160
Pseudo R^2	0.023	0.035	0.075	0.162	0.266	0.341

The table reports the results of the PPML estimation with respect to the $I_{out,i,t}$ indicator. In the first panel we report the estimated coefficients for the variables included in the estimation (see Table 2 for details). The second panel reports the information on the set of fixed effects used in estimation. The last panel shows the number of observations and the Pseudo R^2. The results are obtained via the STATA routine PPMLHDFE. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

churn indicator of at least 32% ($\beta = -0.384$) in the last specification. This result confirms the effect related to the variety of degree programs. Indeed, having more professors in a specific disciplinary area enables institutions to provide students with more and better-tailored courses. Moreover, more professors also imply a lower ratio of students per professor, thus improving courses' quality and the student-teacher relationship. The impact of administrative staff, which is more connected to the quality of services (organization, administration, etc.) that a university could offer, is even larger ($\beta = -1.510$), yielding a reduction of the churn indicator that could reach 77%.

For what concerns the revenues, their growth increases the probability of losing students (being constant regional GDP per capita and taxable income): an increase in university revenues per student is associated with a growth of the churn-out indicator of 20% ($\beta = 0.198$). Indeed, higher revenues can be related to higher enrollment costs, reducing units' attractivity; a share of that revenue is due to public national expenditures, and students' own fees are responsible for a considerable quota.

To better understand the role of degree programs' offering on churn-out flows, Table 4 shows the results regarding the specifications (4) and (6) of Table 3 by separately considering the students that change only the university and those that change also the CUN area.[15]

The results indicate that the effect of the number of degree programs changes depending on the source of variation considered. When the within-university variation is considered (specification (4)), the number of degree programs has a stronger effect in reducing the churn rate of students who change both the university and the CUN area. However, this result changes when we account for the time-invariant characteristics of the CUN areas (specification (6)). In this case, the effect of degree programs' variety is stronger for those who change only the university. Therefore, the negative effect associated with degree programs' supply in Table 3 is mainly driven by those students who seek a specific course within their disciplinary area. This result implies that the variety of the degree programs offered within the various disciplinary areas is crucial to reducing the loss of students due to churn-out behaviors. The more the offer is diversified, the less the students are motivated to seek other alternatives in other institutions.

[15] We thank the anonymous referee for this suggestion.

Table 4
PPML estimation results for the $I_{out,i,t}$ indicator by type of churn.

	Only university		University & CUN area	
	(4)	(6)	(4)	(6)
N of degree programs	−0.0263*	−0.157***	−0.171***	−0.0662***
	(0.0122)	(0.0150)	(0.0123)	(0.0127)
Revenues/N (10k€)	0.282***	0.244**	0.101	0.192**
	(0.0853)	(0.0788)	(0.0784)	(0.0699)
Scholars/N	0.385	0.281	−1.137	−1.347
	(0.994)	(0.954)	(0.867)	(0.708)
Professors in the area/N	−0.271	0.0298	0.151	−0.496***
	(0.225)	(0.266)	(0.102)	(0.138)
Non academics/N	−1.995*	−1.731*	−0.530	−0.795
	(0.859)	(0.809)	(0.795)	(0.639)
Provincial Unemployment	−2.335**	−3.184***	−2.163**	−1.575**
	(0.724)	(0.677)	(0.664)	(0.526)
Taxable income per capita (10k€)	−1.697***	−0.472**	0.438***	−0.864***
	(0.149)	(0.158)	(0.108)	(0.102)
Regional GDP per capita (10k€)	0.895**	0.305	−0.610**	−0.101
	(0.317)	(0.308)	(0.227)	(0.175)
Constant	4.910***	4.428***	5.444***	6.810***
	(0.840)	(0.789)	(0.603)	(0.468)
CUN area FE	No	Yes	No	Yes
University FE	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes
Observations	8151	8151	8135	8135
Pseudo R^2	0.159	0.275	0.170	0.467

The table reports the results of the PPML estimation with respect to the $I_{out,i,t}$ indicator by separately considering the students that change only the university and those that change also the CUN area. In the first panel we report the estimated coefficients for the variables included in the estimation (see Table 2 for details). The second panel reports the information on the set of fixed effects used in estimation. The last panel shows the number of observations and the Pseudo R^2. The results are obtained via the STATA routine PPMLHDFE. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 5 shows the results related to the churn decision within a university, i.e., students who decide to change only the CUN area at the end of their first year but stay in the same university. In that regard, this churning decision is a potential cost for the CUN field and/or the department but not for the university itself. In this case, the effect associated with the number of degree programs changes according to

Table 5
PPML estimation results for the $I_{in,i,t}$ indicator.

	(1)	(2)	(3)	(4)	(5)	(6)
N of degree programs	−0.000282	−0.0680***	−0.0769***	−0.133***	0.193***	0.130***
	(0.00972)	(0.0105)	(0.0105)	(0.0115)	(0.0147)	(0.0141)
Revenues/N (10k€)		−0.586***	−0.616***	−0.0714	−0.263***	0.0236
		(0.0480)	(0.0529)	(0.127)	(0.0474)	(0.113)
Scholars/N		4.237***	4.447***	1.673	3.498***	1.315
		(0.404)	(0.439)	(1.691)	(0.385)	(1.331)
Professors in the area/N		0.603***	0.609***	0.551***	−0.290*	−0.313*
		(0.0988)	(0.104)	(0.0940)	(0.132)	(0.125)
Non academics/N		−1.621***	−1.795***	0.297	−1.659***	0.213
		(0.278)	(0.309)	(1.793)	(0.282)	(1.406)
Provincial Unemployment		−0.964	−0.955	−3.034***	−0.457	−2.144***
		(0.615)	(0.669)	(0.742)	(0.541)	(0.636)
Taxable income per capita (10k€)		1.873***	1.759***	3.199***	0.493***	1.413***
		(0.0964)	(0.119)	(0.214)	(0.107)	(0.238)
Regional GDP per capita (10k€)		−1.096***	−1.235***	−2.113***	−0.718***	−1.290***
		(0.0725)	(0.147)	(0.288)	(0.124)	(0.229)
Constant	3.946***	2.892***	3.480***	3.048***	4.551***	4.414***
	(0.0319)	(0.208)	(0.329)	(0.746)	(0.275)	(0.598)
CUN area FE	No	No	No	No	Yes	Yes
University FE	No	No	No	Yes	No	Yes
Region FE	No	No	Yes	Yes	Yes	Yes
Observations	8160	8160	8160	8061	8160	8061
Pseudo R^2	0.000	0.144	0.168	0.262	0.512	0.567

The table reports the results of the PPML estimation with respect to the $I_{out,i,t}$ indicator. In the first panel we report the estimated coefficients for the variables included in the estimation (see Table 2 for details). The second panel reports the information on the set of fixed effects used in estimation. The last panel shows the number of observations and the Pseudo R^2. The results are obtained via the STATA routine PPMLHDFE. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

the specification. In the first *null* model, which includes only a constant and no other variables, such an effect is not significant, given that all the controls/covariates are missing. As we include more variables in the model, the effect becomes more specific; in the first models, with fewer controls, such an effect is negative, due to missing fixed effects in the model specification. Thus, it becomes positive when all the variables are considered, highlighting a more precise behavior.

Looking at the fourth model, that includes university and regional fixed effects but without specific CUN area effect, the results can be interpreted in a framework of competition between CUN areas in the same university (and, to an essentially smaller extent, among different university locations of the same university). As shown in Table 5, the coefficient for the number of degree programs is $\beta = -0.133$, and it means that if there is a unitary increase in the number of degrees, there is a reduction of 12.45% in the churn rate regarding students that relocate within the same university. Thus, offering more degree programs for the CUN area in the same university reduces this specific churning risk. Universities with more degree programs may provide students with specific courses tailored to their preferences, reducing their need to change degree programs after one year.

In the fifth model, we remove the fixed effects for the university while we introduce the CUN area fixed effects. Here, we deepen the between-university competition phenomenon, given that we are controlling for the CUN area variables. In this case, the coefficient for the N of degree programs is positive: $\beta = 0.193$; thus, for each increase in the number of degree programs, churning risk increases by 21.28%. Increasing the options available seems to increase the chances that units will experience an increase in the churn rate related to students' inter-degree relocation within the same university. This element may be related to the fact that, within the same CUN area, a more varied supply of degree programs may provide courses characterized by many similar exams that can be transferred easily from one CUN area to another. The results also confirm the positive role of having more professors in a given disciplinary area. Controlling for the CUN areas and the universities' characteristics (specifications (5) and (6)), the estimated coefficient is negative: more professors in the area lower the churn indicator by 26.9% ($\beta = -0.313$). As previously mentioned, this result can be linked to the quality improvement of the courses, due to fewer students per professor and more customized degree programs based on students' needs.

Overall, two sets of models indicate that increasing the options of degree programs available for students has twofold effects: churn between universities is reduced, and churning within universities increases. Thus, students are likely kept in the same university (in this regard, it is a *negated risk*) as more programs are offered. Furthermore, for a student, the option to change the CUN area (i.e., degree program) could be considered an improvement, given that it occurred only after one year, and such a student is ideally on time in changing academic career and being successful. In contrast, being kept in a degree program not suited for such a student will likely lead to poor academic proficiency or, in the worst scenario, even a complete dropout from the university system. Thus, the idea that each churn has to be conceived as a risk for a single student is, of course, a simplification; it can even be an opportunity to change their path in time. Therefore, being able to offer a wider range of degree programs can be a threat to specific CUN areas. However, it may help universities retain their students while attracting new ones from competitor universities.

6. Conclusion and policy implication

In Italy, as elsewhere in the world, the quality of the educational and training systems is a strategic element from a personal point of view and to support social progress and economic development of the society as a whole. Education contributes to human capital development, research, and, in developing countries, even nation-building itself [75]. Reducing the dropout and churn phenomena are certainly issues to be addressed to improve the educational system's quality. They have been considered in the socio-economic literature at the same time as (i) a waste of time and public and private financial resources, and (ii) an inevitable step for some students who entered college with inadequate knowledge of themselves [52]. While for each individual student, the churn decision could be an even necessary pathway to increase educational success after choosing a more suited degree program for himself/herself, when dealing with the analyses of higher education as a system, it is without any doubt an overall negative phenomenon due to its cost (both economic and social).

In this respect, this paper focuses on the critical issues related to the churn phenomenon in Italy, which has become even more evident in a quasi-market context where universities compete to attract students into their degree programs [2,5,16]. Within this context, compounded by the evaluation criteria of Italian universities, evaluating the elements that affect students' churn risk has become crucial for policymakers and institutions to implement effective strategies to support student success. Indeed, by early identifying at-risk students, universities can reduce the tangible and intangible costs attributable to churn risk in terms of, for example, the organization of space and facilities, which are adapted and equipped according to the number of enrollments, and compensate for the number of students who have dropped out or churned with new enrollments [59]. The loss of a year of study has several implications for the student's future, such as economic costs (fees paid), delay in degree achievement, and late entry into the job market. Within this framework, it is evident that managing the risk of students' career disruption is a multidimensional issue influenced by different factors related to individual characteristics (e.g., academic preparedness, family background, mismatched expectations). Thus, academic challenges require a comprehensive approach. Indeed, only by recognizing and pointing out these challenges can stakeholders work towards improving retention rates and ensuring that more students successfully complete their higher education programs.

To better understand the elements that may help universities in reducing the churn phenomenon, this research has investigated the relationship between universities' attributes and the risk of churn associated with inter-degree and inter-university relocations by exploiting the administrative data on the population of students enrolled in the Italian higher education system between 2010 and 2018 embedded in a data-integration approach. In particular, the empirical approach has been based on the application of the PPML estimator to account for the presence of zeros in dependent variables (due to the absence of churners in the units) and on the inclusion of several control variables and fixed effects to account for the role played by the characteristics of universities, CUN areas, and university's hosting areas.

The results highlight that the variety and the dimension of universities' services, also in the form of human resources (researchers, professors, and non-strictly academic figures, such as administrative staff), contribute to reducing students' churn flows and are positively associated with universities' retention rates. This result may be related to the opportunity for institutions to diversify their educational offer by providing more and better-tailored courses to meet students' preferences, even within each CUN area. In this respect, the number of degree programs and the number of professors in the disciplinary area are essential for improving institutions' quality and matching students' preferences. Moving from this result and considering the aging population of Italy that will have negative effects on the number of students enrolled in universities [76], it might be beneficial, while keeping the public costs of education constant, to keep also the number of professors constant to improve the ratio between students and professors and enhance the quality of the degree programs by providing a more customized offering.

Furthermore, from the analysis of within-university relocations, it is clear that increasing the available programs could also help students rethink their academic path after one year while staying in the same institution. In this respect, a good higher education orientation is critical to reducing students' career path disruption at the university [77]. Indeed, universities have promoted several measures to facilitate student retention, thereby avoiding degree program switching that could lead to both study delay and negative evaluation for universities. In order to guide students' choices based on their knowledge and skills, the Italian National Recovery and Resilience Plan (Piano Nazionale di Ripresa e Resilienza — PNRR), which is part of the Next Generation EU recovery project, has defined some measures related to students' orientation in school-university transition in Mission 4 (Education and Research), Investment 1.6, aiming at qualifying the education system and increase the number of graduates by facilitating and encouraging the transition from upper secondary school to university and, at the same time, addressing the problem of university dropouts in later years with targeted guidance and mentoring actions. Regarding this aspect, universities have already introduced some improvements in the orientation programs, such as the activities implemented under the Italian Ministerial Decree 934/22 — Criteria for allocation of resources and methods of implementation of projects related to "Active guidance in school-university transition" (PNRR) aimed to lead students toward a much more reasoned and well-informed choice of the degree program.

This contribution adopted a systematic analytical approach that considers variables observed at the macro level of universities and CUN disciplinary areas. Therefore, the chosen approach has the drawback of not considering the perspective of single students who are faced with the difficult decision of whether to continue their studies at university. On the one hand, the proposed approach has allowed us to link the phenomenon of churn-out risk to the structural characteristics of the university of enrollment and the reference context. On the other hand, the point of view of churning students has not been considered. However, the detection of individual aspects through, for example, the creation of an ad hoc survey, could have enriched the discussion by suggesting appropriate improvement measures able to offer university governance valuable tools to prevent the risk of losing students in the transition from one year to the successive one. Investigate factors such as students' satisfaction with respect to the course of study, the teaching provided, and the services offered by the university; students' perception of social support, as well as expectations about their professional and personal future, together with other factors expressly linked to the work context (for example, job offers after graduation), are key factors to consider for further development of researches on the topic of university dropout and churn in general. Some of these features could be linked to the environmental variables that we included in the models, acting as proxies, but the individual point of view has not been included explicitly in the analyses. A challenge could be to perform altogether a micro-level survey embedded in a systematic analysis that points instead to the macro-level, to obtain an exhaustive picture of the churning problem for the Italian higher education system.

CRediT authorship contribution statement

Cristian Usala: Conceptualization, Data curation, Formal analysis, Methodology, Writing – review & editing. **Ilaria Primerano:** Conceptualization, Data curation, Investigation, Writing – original draft, Writing – review & editing. **Francesco Santelli:** Conceptualization, Data curation, Investigation, Writing – original draft, Writing – review & editing. **Giancarlo Ragozini:** Conceptualization, Resources, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgments

Financial support under the National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 1.1, Call for tender No. 104 published on 2.2.2022 by the Italian Ministry of University and Research (MUR), funded by the European Union – NextGenerationEU– Project Title From high school to university: Assessing peers' influence in educational inequalities and performances – CUP F53D23006150006 - Grant Assignment Decree No. 1060 adopted on 07/17/2023 by the Italian Ministry of Ministry of University and Research (MUR)

We thank Claudio Deiana for helpful advice and clarifications on the Poisson Pseudo Maximum Likelihood estimator and its application.

Data analysis manipulation

The data used in this study have been processed in accordance with the RESEARCH PROTOCOL FOR THE STUDY 'From high school to the job placement: analysis of university careers and university mobility from Southern to Northern Italy' among the Ministry of University and Research, the Ministry of Education and Merit, the University of Palermo as the lead institution, and the INVALSI Institute. The reference researcher is Giancarlo Ragozini.

References

[1] Helsloot I, Jong W. Risk management in higher education and research in the Netherlands. J Contingencies Crisis Manag 2006;14(3):142–59. http://dx.doi.org/10.1111/j.1468-5973.2006.00490.x, URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-5973.2006.00490.x.

[2] Lumino R, Gambardella D, Grimaldi E. The evaluation turn in the higher education system: Lessons from Italy. J Educ Administr Hist 2017;49(2):87–107.

[3] Miller K, McAdam M, McAdam R. The changing university business model: A stakeholder perspective. R D Manag 2014;44(3):265–87.

[4] Deem R, Brehony KJ. Management as ideology: The case of 'new managerialism'in higher education. Oxf Rev Educ 2005;31(2):217–35.

[5] Marginson S. Dynamics of national and global competition in higher education. Higher Educ 2006;52:1–39.

[6] Rothschild M, White LJ. The analytics of the pricing of higher education and other services in which the customers are inputs. J Polit Econ 1995;103(3):573–86.

[7] Cattaneo M, Malighetti P, Meoli M, Paleari S. University spatial competition for students: The Italian case. Reg Stud 2017;51(5):750–64.

[8] Ubi J, Liiv I. A review of student Churn in the light of theories on business relationships. In: Educational data mining 2010. ERIC; 2010.

[9] Santos Silva JMC, Tenreyro S. The Log of Gravity. Rev Econ Stat 2006;88(4):641–58. http://dx.doi.org/10.1162/rest.88.4.641, arXiv:https://direct.mit.edu/rest/article-pdf/88/4/641/1982805/rest.88.4.641.pdf.

[10] OECD. Education policy analysis 2006: Focus on higher education. Organisation for Economic Co-operation and Development; 2006.

[11] Curaj A, Deca L, Pricopie R. European higher education area: Challenges for a new decade. Springer Nature; 2020.

[12] Agasisti T, Pérez-Esparrells C. Comparing efficiency in a cross-country perspective: The case of Italian and Spanish state universities. Higher Educ 2010;59:85–103.

[13] Ezza A, Pischedda G, Marinò L. Performance-based funding in public competition. lights and shadows in the Italian higher education system. J Int Bus Econ 2017;17(2):5–22.

[14] Ezza A, Fadda N, Pischedda G, Marinò L. Il" grande gap": Gli effetti del performance budgeting sulle politiche di reclutamento delle università Italiane. Manag Control 2019;(2019/2).

[15] Dal Molin M, Turri M, Agasisti T. New public management reforms in the Italian universities: Managerial tools, accountability mechanisms or simply compliance? Int J Public Administr 2017;40(3):256–69.

[16] Agasisti T, Catalano G. Governance models of university systems—towards quasi-markets? tendencies and perspectives: A European comparison. J High Educ Policy Manag 2006;28(3):245–62.

[17] Geuna A, Rossi F. The university and the economy: pathways to growth and economic development. Edward Elgar Publishing; 2015.

[18] Dotti NF, Fratesi U, Lenzi C, Percoco M. Local Labour Markets and the Interregional Mobility of Italian University Students. Spat Econ Anal 2013;8(4):443–68. http://dx.doi.org/10.1080/17421772.2013.833342.

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

[20] Tosi F, Impicciatore R, Rettaroli R. Individual skills and student mobility in Italy: A regional perspective. Reg Stud 2019;53(8):1099–111.

[21] Valero A, Van Reenen J. The economic impact of universities: Evidence from across the globe. Econ Educ Rev 2019;68:53–67. http://dx.doi.org/10.1016/J.ECONEDUREV.2018.09.001.

[22] Fadda N, Marinò L, Pischedda G, Ezza A. The effect of performance-oriented funding in higher education: Evidence from the staff recruitment budget in Italian higher education. Higher Educ 2022;83(5):1003–19.

[23] Contini D, Cugnata F, Scagni A. Social selection in higher education. Enrolment, dropout and timely degree attainment in Italy. Higher Educ 2018;75(5):785–808.

[24] Romito M, Pilutti S, Contini D. Why do students leave university? Qualitative research at an Italian higher education institution. Eur J Educ 2020;55(3):456–70.

[25] La Rocca M, Niglio M, Restaino M. Predicting university student' churn risk. In: Book of short papers IES 2022 innovation & society 5.0: statistical and economic methodologies for quality assessment. PKE srl; 2022, p. 60–2.

[26] Primerano I, Giordano G, et al. An exploratory strategy for analyzing students' mobility data. In: Book of short papers 11th international conference IES 2023 statistical methods for evaluation and quality: techniques, technologies and trends (t3). Edizioni BACME Srls; 2023, p. 65–70.

[27] Pitzalis M, Porcu M. Passaggio a Nord. Come si ristruttura il campo universitario italiano? Scuola democratica 2015;3:711–22. http://dx.doi.org/10.12828/82088.

[28] Attanasio M, Enea M. La mobilitá degli studenti universitari nell'ultimo decennio in Italia. In: Santis GD, Pirani E, Porcu M, editors. Rapporto sulla popolazione. l'istruzione in Italia. Bologna: Il Mulino; 2019, p. 43–58.

[29] Lombardi G, Ghellini G. The effect of grading policies on Italian universities' attractiveness: A conditional multinomial logit approach. Electron J Appl Stat Anal 2019;12(4):801–25.

[30] Attanasio M. Students' university mobility patterns in Europe: An introduction. Genus 2022;78(17). http://dx.doi.org/10.1186/s41118-022-00164-8.

[31] Restaino M, Vitale MP, Primerano I. Analysing international student mobility flows in higher education: A comparative study on European countries. Soc Ind Res 2020;149:947–65.

[32] Breznik K, Skrbinjek V. Erasmus student mobility flows. Eur J Educ 2020;55(1):105–17.

[33] Boscaino G, Genova V, et al. Pull factors for university students' mobility: A gravity model approach. In: Book of short papers-SIS 2020. Pearson; 2020, p. 73–8.

[34] Columbu S, Porcu M, Sulis I. 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 2021;74:100926.

[35] Enea M. 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, editors. Studies in theoretical and applied statistics. SIS 2016. Springer Proceedings in Mathematics and Statistics, vol 227, Cham: Springer; 2018, p. 239–49. http://dx.doi.org/10.1007/978-3-319-73906-9_22.

[36] D'Agostino A, Ghellini G, Longobardi S. Out-migration of university enrolment: The mobility behaviour of Italian students. Int J Manpower 2019;40:56–72. http://dx.doi.org/10.1108/IJM-07-2017-0169.

[37] Genova VG, Tumminello M, Enea M, Aiello F, Attanasio M. Student mobility in higher education: Sicilian outflow network and chain migrations. Electron J App Stat Anal 2019;12(4):774–800. http://dx.doi.org/10.1285/I20705948V12N4P774.

[38] Santelli F, Scolorato C, Ragozini G. On the determinants of student mobility in an interregional perspective: A focus on Campania region. Ital J Appl Stat 2019;31(1):119–42. http://dx.doi.org/10.26398/IJAS.0031-007.

[39] Bacci S, Bertaccini B. Assessment of the university reputation through the analysis of the student mobility. Soc Ind Res 2021;156:363–88.

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

[41] Santelli F, Di Credico G, Di Caterina C, et al. High school proficiency of future university students: An analysis based on INVALSI data. In: Book of short papers 11th international conference IES 2023 statistical methods for evaluation and quality: techniques, technologies and trends (t3). Edizioni" Il Viandante"; 2023, p. 560–5.

[42] Di Credico G, Di Caterina C, Santelli F, et al. Streamlined variational inference for modeling Italian educational data. In: Proceedings of the statistics and data science conference. PaviaUniversityPress; 2023, p. 102–7.

[43] Columbu S, Porcu M, Primerano I, Sulis I, Vitale MP, et al. Correction to: Analysing the determinants of Italian university student mobility pathways,(genus,(2021) 77, 1,(34), 10.1186/s41118-021-00146-2). Genus 2022;78(1).

[44] Primerano I, Santelli F, Usala C. A multiplex network approach for analyzing university students' mobility flows. In: Studies in theoretical and applied statistics: SIS 2021, pisa, Italy, June 21–25. Springer; 2023, p. 75–89.

[45] Barbieri B, Porcu M, Salaris L, Sulis I, Tedesco N, Usala C. University dropout and Churn in Italy: An analysis over time. In: SIS 2022-51th scientific meeting of th Italian statistical society-book of the short papers. Pearson; 2022, p. 1734–9.

[46] Primerano I, Santelli F, Usala C, Ragozini G. Exploiting students' inter-degree relocations to assess Italian universities' attractiveness. In: Book of abstract 9^{rh} international conference on risk analysis. 2022, p. 60–2.

[47] Delogu M, Lagravinese R, Paolini D, Resce G. Predicting dropout from higher education: evidence from italy. Econ Mod 2024;130:106583. http://dx.doi.org/10.1016/j.econmod.2023.106583.

[48] Columbu S, Porcu M, Primerano I, Sulis I, Vitale MP. Geography of Italian student mobility: A network analysis approach. Socio-Econ Plan Sci 2021;73:100918.

[49] Figini S, De Quarti E, Giudici P. Chrun risk mitigation models for student's behavior. Electron J Appl Stat Anal 2009;2(1):37–57.

[50] Sarra A, Fontanella L, Di Zio S. Identifying students at risk of academic failure within the educational data mining framework. Soc Ind Res 2019;146:41–60.

[51] Impicciatore R, Tosi F. Student mobility in Italy: The increasing role of family background during the expansion of higher education supply. Res Soc Strat Mobil 2019;62:100409. http://dx.doi.org/10.1016/j.rssm.2019.100409.

[52] Aina C, Baici E, Casalone G, Pastore F. The determinants of university dropout: A review of the socio-economic literature. Socio-Econ Plan Sci 2022;79:101102.
[53] Usala C, Porcu M, Sulis I. The high school effect on students' mobility choices. Stat Methods Appl 2023;32(4):1259–93.
[54] Bratti M, Verzillo S. The 'gravity' of quality: Research quality and the attractiveness of universities in Italy. Reg Stud 2019;53(10):1385–96. http://dx.doi.org/10.1080/00343404.2019.1566701.
[55] Pigini C, Staffolani S. Beyond participation: Do the cost and quality of higher education shape the enrollment composition? The case of Italy. Higher Educ 2016;71(1):119–42. http://dx.doi.org/10.1007/s10734-015-9892-8.
[56] Giambona F, Porcu M, Sulis I. Students mobility: Assessing the determinants of attractiveness across competing territorial areas. Soc Indic Res 2017;133(3):1105–32. http://dx.doi.org/10.1007/s11205-016-1407-1.
[57] Whitlock JL. Using data science and predictive analytics to understand 4-year university student churn [Ph.D. thesis], East Tennessee State University; 2018.
[58] Supangat S, Saringat MZB, Kusnanto G, Andrianto A. Churn prediction on higher education data with fuzzy logic algorithm. SISFORMA 2021;8(1):22–9.
[59] Djulovic A, Li D. Towards freshman retention prediction: A comparative study. Int J Inform Educ Technol 2013;3(5):494–500.
[60] Daraio C, Glänzel W. Grand challenges in data integration—State of the art and future perspectives: An introduction. Scientometrics 2016;108(1):391–400.
[61] [database] Database MOBYSU IT. Mobilitá 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.
[62] Santos Silva J, Tenreyro S. Further simulation evidence on the performance of the Poisson pseudo-maximum likelihood estimator. Econom Lett 2011;112(2):220–2. http://dx.doi.org/10.1016/j.econlet.2011.05.008, URL https://www.sciencedirect.com/science/article/pii/S0165176511001741.
[63] Manning WG, Mullahy J. Estimating log models: To transform or not to transform? J Health Econ 2001;20(4):461–94. http://dx.doi.org/10.1016/S0167-6296(01)00086-8, URL https://www.sciencedirect.com/science/article/pii/S0167629601000868.
[64] Powell D, Seabury S. Medical care spending and labor market outcomes: Evidence from workers' compensation reforms. Amer Econ Rev 2018;108(10):2995–3027. http://dx.doi.org/10.1257/aer.20150912, URL https://www.aeaweb.org/articles?id=10.1257/aer.20150912.
[65] Wooldridge JM. Econometric analysis of cross section and panel data. MIT Press; 2010.
[66] Gourieroux C, Monfort A, Trognon A. Pseudo maximum likelihood methods: Applications to Poisson models. Econometrica 1984;52(3):701–20, URL http://www.jstor.org/stable/1913472.
[67] Correia S, Guimarães P, Zylkin T. Fast Poisson estimation with high-dimensional fixed effects. Stata J 2020;20(1):95–115, arXiv:10.1177/1536867X20909691.
[68] Cragg JG. Some statistical models for limited dependent variables with application to the demand for durable goods. Econometrica 1971;39(5):829–44, URL http://www.jstor.org/stable/1909582.
[69] Mullahy J. Specification and testing of some modified count data models. J Econometrics 1986;33(3):341–65. http://dx.doi.org/10.1016/0304-4076(86)90002-3, URL https://www.sciencedirect.com/science/article/pii/0304407686900023.
[70] Santos Silva J, Tenreyro S. The log of gravity at 15. School of economics discussion papers 0121, School of Economics, University of Surrey; 2021, URL https://ideas.repec.org/p/sur/surrec/0121.html.
[71] Ederington J, Paraschiv M, Zanardi M. The short and long-run effects of international environmental agreements on trade, no. 242514732. Lancaster University Management School, Economics Department; 2018, URL https://ideas.repec.org/p/lan/wpaper/242514732.html.
[72] Bove V, Deiana C, Nisticò R. Global arms trade and oil dependence*. J Law Econ Organ 2018;34(2):272–99. http://dx.doi.org/10.1093/jleo/ewy007.
[73] Colagrossi M, Deiana C, Dragone D, Geraci A, Giua L, Iori E. Intimate partner violence and help-seeking: The role of femicide news. J Health Econ 2023;87:102722. http://dx.doi.org/10.1016/j.jhealeco.2022.102722, URL https://www.sciencedirect.com/science/article/pii/S0167629622001369.
[74] Larch M, Wanner J, Yotov YV, Zylkin T. Currency unions and trade: A PPML re-assessment with high-dimensional fixed effects. Oxf Bull Econ Stat 2019;81(3):487–510, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/obes.12283.
[75] Lawal RO, Akinyemi IA, Gbenu JP. Quality assurance practices: Veritable tools for goal achievement in tertiary educational institutions in Nigeria. Afr J Educ Manag 2021;22(1):235–53.
[76] Billari F. Domani eóggi': costruire il futuro con le lenti della demografia. EGEA; 2024.
[77] Cavalletti B, Corsi M, Persico L, di Bella E. Public university orientation for high-school students. a quasi-experimental assessment of the efficiency gains from nudging better career choices. Socio-Econ Plan Sci 2021;73:100945.

Cristian Usala is a Ph.D. Assistant Professor (Non Tenured/RTDa) in Social Statistics at the University of Cagliari. His research interests include discrete choice analysis, education, peer effects, energy communities.

Ilaria Primerano is technologist at the Institute for Research on Population and Social Policies of the National Research Council. Her research interests concern the analysis of complex data structures using Multidimensional Data Analysis and Social Network Analysis with applications in economic and social fields, with particular attention to the evaluation of university teaching, national and international student mobility, university dropouts, social support and network society.

Francesco Santelli is Ph.D.Assistant Professor (Non Tenured/RTDa) at University of Trieste, Department of Economics, Business, Mathematics and Statistics (DEAMS). His research topics are within the Social Statistics framework; Social Media Data, Social Network Analysis, Statistical Textual Analysis, Mobility Patterns, statistical tools for Higher Education and Social Support Networks.

Giancarlo Ragozini is Full Professor of Social Statistics at the Department of Political Science, University of Naples Federico II. His research expertise covers the statistical methods for social network analysis, computational statistics, and multivariate methods for data analysis. He is author of several publications in high impact international scientific journals like Social Network, Network Science, Computational Statistics and Data Analysis, Cliometrica. He participated and managed several Italian and European projects on official statistics, youth policies, evaluation of public policies, network analysis applications. Nowadays he is board member of the Italian Association of Population Studies. He organized several national and international meeting, especially on social network topics.