

University choice and the attractiveness of the study area. Insights from an analysis based on generalized mixed-effect models

Silvia Columbu, Mariano Porcu and Isabella Sulis

Abstract In this work we investigate upon the determinants of students' choices to attend bachelor degree studies outside the region of residence using information provided by the Italian National Student Archive (NSA) on a cohort of students enrolled for the first time at the university in a.y. 2014/15. The aim of the analysis is twofold: (i) to suggest value-added measures of university reputation and (ii) to assess and split the role played by the field of study in determining the power to attract students from other regions. To do this, a three-level logistic regression model to deal with cross-classified units has been adopted for modelling the probability that freshmen in a given university are mover students (instead that stayer) as a function of their socio-demographic characteristics, territorial area information and other sources of heterogeneity which concern both the field of specialization and the university reputation.

Abstract In questo lavoro, considerando i dati dell'Anagrafe Nazionale Studenti (ANS) per la coorte di immatricolati nell'a.a. 2014/15, si studiano i fattori che spingono gli studenti italiani a lasciare la propria regione di residenza per intraprendere gli studi universitari. L'analisi riportata si pone due obiettivi principali: (i) suggerire delle misure di valore aggiunto della reputazione delle università e (ii) individuare l'influenza dell'indirizzo di studi nella capacità di attrarre studenti da altre regioni. Si considera un modello di regressione logistica a tre livelli, che parametrizza una struttura cross-classified, per modellizzare la probabilità che uno studente al primo anno sia fuori sede in funzione delle sue caratteristiche socio-demografiche, delle informazioni territoriali, e di altre fonti di eterogeneità relative all'area di specializzazione scelta e alla reputazione dell'università.

Key words: multilevel models, value-added, student mobility, university reputation

1 Introduction

In the last years there has been an increasing request to evaluate the effectiveness of the tertiary education institutions; these evaluations have been used to support and enhance universities and curricula which satisfy specific quality standards. The development of the assessment policies in tertiary education has determined a huge contraction of curricula supplied in many universities and a redefinition of the shape of the educational offer. Among the indicators used by the government to determine the share of the yearly public financial provisions transferred to public universities are considered, also, the job placement of graduates and the universities' capability to attract students from other territories. However, students' mobility choices are related to students' socio-economic conditions and to the economic peculiarities of both the geographical areas of students' provenance and the place where the universities are located, as well as to the university reputation and effectiveness [6] [2] [8]. Recent empirical studies on student mobility in Italy, highlight the presence of a North-South divide in the way university students make their choices [7][8]. Specifically, the distribution of university students, with respect to the region where they reside (origin) and the region where they attend the university studies (destination), shows that about 1 out of 4 freshmen students in Italy move from the South to the North or Center regions, while very few move from North-Center to South.

Many authors have addressed phenomenon of intra-and international students mobility. In an economic perspective, Dotti et al. (2013) [6] state that the migration of students in Italy can be seen as a reaction to low employment rates in the provinces of origin, and is directed mainly to the provinces with better job opportunities. D'Agostino *et al.* (2016) [5] discuss the geographical dimension of students mobility in the Italian case. Enea (2016) [7] analyses the student Italian mobility case in the transition from 1st level degree to master, highlighting that students with the best educational background have the high risk to move to another university for their master studies and that, among these, about 75% chooses a northern University. Suhonen (2014) [11] proves that in the Finnish university system the distance effect is highly conditioned to the choice of specific fields of study. The geographical dimension has also been investigated by Cattaneo et al (2016) [3], who focus on the role played by transport accessibility in determining students' choices and prove that the easiness to move is a factor of attractiveness of Italian universities. There is evidence that the migration phenomenon can be linked to the quality of university research and teaching in the Italian context ([4][2](2016) [10]). In an international framework, Beine et al. (2014) [1] identified that there exists a network effect, that contributes to spread the reputation of a university and attract new students. Giambona, Porcu & Sulis (2016) [8] highlight that the determinants of students' flows among geographical areas seem more related to the opportunities offered by the destination areas than to the characteristics of their universities in terms of facilities and services.

Moving from this framework, we define *mover* those students who enroll in universities that are not located in the region where they reside and *stayer* those who come from the same region. Since the raw rate of *movers* enrolled in a university

does not allow to determine which differences in university attractiveness are related to students' characteristics and/or socio-economic background of the territorial areas of the universities and which differences are related to the vocation of the universities in terms of their education offering and/or to better education opportunities and higher level of competencies, we advance a micro analysis of the determinants of students' university choices.

The aim of the analysis is to assess and split the roles played by the field of study and by the university in determining the power to attract students from other regions. To do this, a three-level logistic regression model has been adopted for modelling the probability that freshmen in a given university are *mover* students (instead that *stayer*) as a function of their socio-demographic characteristics, territorial area information and other sources of heterogeneity of the universities; this last one concern both the field of specialization and the university reputation.

2 Data description

We use the data from Italian National Students Administrative Archive (*NSA*) provided by the Ministry of Education and Research (*MIUR*) from a cohort of students' enrolled for the first time, at a public or private university located in the national territory, in the a.y. 2014/15. We consider only records of students enrolled in traditional Bachelor or in Single-Cycle (Bachelors+Masters) Degree Programs, excluding those related to e-learning degree programs (i.e. the 10 Italian universities that offer only e-learning degree programs) or to foreign students. The analysis has been bounded to 229,813 observations, belonging to 80 (61 public and 29 private) universities that provide traditional tertiary education programs in Italy. The *NSA* collects information on students' socio-demographic characteristics, information on previous studies, university curricula (i.e. degree program and university where they enroll) and proficiency in the university studies. According to their place of origin, students are classified in 106 provinces and 20 regions. Only 99 out of 106 provinces host at least one university.

In Italy, we observe at least one university per region, thus we could advance the hypothesis that all students considered in the analysis are *free movers* [8]; namely, they all had the opportunity to attend their university studies in the region where they reside but decided to move to another. Overall, 56496 *mover* students have been identified among the subset of 229,813 students. Descriptive statistics clearly show that the propensity to be a *mover* varies across universities, the kind of degree program attended, the macro-area and the region of origin and destination. The research combines personal details on students, provided by the *NSA*, with socio-economic indicators of the provinces where the universities are located, provided by the Italian National Institute of Statistics (*ISTAT*), and other information on degree programs and universities effectiveness and reputation, provided by the surveys on Graduates' Profile and Graduates' Employment condition (carried out by the *AlmaLaurea Consortium*). Therefore, the following covariates have been considered

to study the determinants of students' mobility choices: a) gender (GENDER), age at enrollment (AGE), the region of residence (REGION) and the information related to high school background- such as final grade obtained at the end of high school (GRADE) and the kind of high school attained (HIGHSCHOOL) - at student level; b) at province level the youth unemployment rate (25-34 years old) (YUNEMPLOYMENT) and a normalized indicator of gross value added per capita (NGVA) have been used to contextualize differences in the socio-economic conditions between students' provinces of residence; moreover, the information on the number of different degree programs (NCOURSES) supplied by the universities located in the province of residence has been used as a proxy of the dimension of the tertiary education supply; c) at degree program level, (DPL) the information on graduates' rate of employment one year after graduation (EmployAfterGrad-DPL), on their average wage (WageAfterGrad-DPL) and on their self-reported perception on several aspects of the universities studies (Satisfaction -DPL), have been considered to contextualize differences in degree programs effectiveness and reputation. The information on the percentage of students with both parents graduated (GradParents -DPL) has been adopted to take into account the heterogeneity of socio-economic conditions of students belonging to different degree programs. Finally (d) at the university level the same indicators used at degree program level and described at point (c) (EmployAfterGrad-UL; Satisfaction -UL; GradParents- UL) have been adopted to contextualize differences in the characteristics of the universities. The macro-area where the university is located has been used to assess the effect of geographical components.

3 Modeling approach

Let us indicate with Y_{ijg} an indicator variable which assumes value 1 if student i ($i = 1, \dots, n$) from degree program j ($j = 1, \dots, J$) of the university g ($g = 1, \dots, G$) resides in the region where the university is located (is a *mover*) and 0 if she/he resides in the same region (is a *stayer*). The probability to be a mover is modeled using a logistic function as follows

$$\text{logit}[\gamma_{ijg}] = \tau + \mathbf{X}_{ijg}^T \boldsymbol{\beta} + \mathbf{Z}_j^T \boldsymbol{\gamma} + \mathbf{U}_g^T \boldsymbol{\delta} + \theta_j + \boldsymbol{\lambda}_g^T \mathbf{v}_g \quad (1)$$

where $\gamma_{ijg} = \pi_{ijg}$ and \mathbf{X}_{ijg} is a vector of individual covariates (Level-1), \mathbf{Z}_j is a vector of covariates at degree program level j (Level-2), and \mathbf{U}_g is a vector of covariates at university level (Level-3). $\theta_j \sim N(0, \sigma_\theta^2)$ is a random term shared by observations related to the same degree program, whereas $\mathbf{v}_g \sim MVN(0, \Sigma_v)$ enables to account for heteroskedasticity between universities due to geographical components (e.g between macro-area propensity to attract students). The presence of both random components enables to split the between university variability from the between degree program variability in the propensity to attract students. $\boldsymbol{\lambda}_g$ is a vector which enables to identify the macro-area where the university is located (e.g

considering three geographical macro areas, λ_g for student i assumes value 1 only for the macro-area where the university is located). The expected posterior predictions of the random terms ($\hat{\theta}_j$ and $\hat{\nu}_g$) enable to make inference about degree programs (i.e. degree program-university combination) and universities attractiveness. The cross-classified model described on (1) (where students are clustered in degree programs (Level-2) that belong to different universities (Level-3)) has been made hierarchical by considering at Level-2 degree program-university combinations. Thus, the variability in the propensity to attract students is split in two components: the between-university variability and the between-degree program within-university variability, allowing the divergences in $\hat{\theta}_{jg}$ to capture the gap in the attractiveness of degree program j belonging to the university g from the overall university attractiveness parameter $\hat{\nu}_g$. The model has been estimated with the `runmlwin` routine which calls MLwiN scripts from Stata by adopting Monte Carlo Markov Chain algorithm (Leckie & Charlton (2013)) [9].

4 Model Results

A model building strategy has been carried out to select relevant predictors at each level of analysis. Starting from the null model, relevant assumptions for the definition of the model structure have been tested. The main findings suggest that the variance of the random terms across geographical areas is homoskedastic. Thus the random component ν_g has been considered univariate. Results depicted in Table 1 show that `AGE` and `HIGHSCHOOL` attended influence the probability to be a *mover*. Namely, it is higher for older students who attended a classical or scientific lyceum and come from southern regions (with the exception of Aosta Valley) such as Basilicata, Calabria, and Apulia, while it is lower for younger students who attended a professional school and come from northern and center regions such as Emilia Romagna, Tuscany and Lombardy. *Movers* come mainly from provinces with disadvantage economic conditions, as it is shown by the coefficient of both socio-economic indicators `YUNEMPLOYMENT` and `NGVA` addressed to depict the socio-economic framework. The dimension of the tertiary education supply in the origin provinces does not seem to be a deterrent to students' propensity to move. Looking at the geographical macro-areas where the universities are located, we can observe that *movers* prefer to study in universities of the north or center of Italy. The employment rate one year after the graduation in a given university seems to be a pull factor of mobility. The characteristics of the degree program highlight that the propensity to be a *mover* is higher in the degree programs attended by students coming from more educated families (`GradParents -DPL`), and lower in those where the perceived student satisfaction towards the university studies is higher (`Satisfaction-DPL`). The variances of the random terms show that about 70% of the variability in the propensity to be a *mover* is explained at University level, whereas differences between degree programs within each university, even if significant, have a lower effect ($\sigma_\nu^2 = 2.84$ vs $\sigma_\theta^2 = 0.335$).

Table 1 Multilevel logistic regression estimates

Fixed Effects	Coeff.	Std. Error	p-value	Fixed Effects	Coeff.	Std. Error	p-value
Students' characteristics				Characteristics of the Territorial areas of origin			
GENDER (Baseline=Male)				REGION (Baseline=ABRUZZO)			
Female	0.015	0.0151		BASILICATA	5.833	0.0973	***
AGE	0.037	0.0019	***	CALABRIA	4.417	0.0948	***
HIGHSCHOOL(Baseline= Other Lyceum)				AOSTA VALLEY	3.427	0.155	***
Classic Lyceum	0.305	0.0244	***	APULIA	2.747	0.0598	***
Foreign School	0.072	0.0486		SICILY	2.273	0.0705	***
Professional School	-0.141	0.0345	***	MOLISE	1.933	0.0802	***
Scientific Lyceum	0.125	0.0205	***	SARDINIA	1.882	0.0931	***
Technical School	-0.102	0.0244	***	CAMPANIA	0.972	0.0565	***
				LIGURIA	0.660	0.0647	***
GRADE	0.006	0.0006	***	UMBRIA	0.228	0.0693	***
University characteristics				TRENTINO ALTO ADIGE			
MACRO AREA(Baseline=CENTER)				PIEDMONT			
NORTH	0.206	0.117	*	MARCHE	-0.576	0.0514	***
SOUTH	-2.525	0.230	***	FRIULI VENEZIA GIULIA	-0.681	0.0699	***
EmployAfterGrad-UL	0.017	0.002	***	VENETO	-0.767	0.0545	***
Degree Program characteristics				LAZIO			
GradParents-DPL	0.010	0.004	**	LOMBARDY	-1.548	0.0488	***
Satisfaction-DPL	-0.020	0.003	***	TUSCANY	-1.637	0.0564	***
EmployAfterGrad-DPL	0.0000766	0.001		EMILIA ROMAGNA	-2.172	0.0525	***
WageAfterGrad-DPL	-0.0000307	0.000		UNEMPLOYMENT	0.011	0.0023	***
				NGVA	-0.024	0.000728	***
				NCOURSES	-0.000584	0.0005	
Random effects				Goodness of fit			
	Sd	SE	p-value	DEVIANCE			
University	2.840	0.516	***	DIC	159711.78		
Degree (by university)	0.335	0.018	***		160820.64		

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

References

1. Beine, M., Noe, R., Ragot, L.: Determinants of the international mobility of students. *Economics of Education Review*, 41, 40-54 (2014)
2. Bratti, M. and Verzillo, S., The 'Gravity' of Quality: Research Quality and Universities' Attractiveness in Italy. IZA Discussion Paper No. 11026 (2017)
3. Cattaneo, M., Malighetti, P., Paleari, S., Redondi, R.: The role of the air transport service in interregional long-distance students' mobility in Italy. *Transportation Research Part A*, 93, 66-82 (2016)
4. Ciriaci, D.: Does University Quality Influence the Interregional Mobility of Students and Graduates? The Case of Italy, *Regional Studies*, 48(10), 1592-1608 (2014)
5. D'agostino A., Ghellini G., Longobardi S. : University mobility at enrollment: geographical disparities in Italy. *Atti della XLVIII Riunione Scientifica SIS 2016*, Salerno 8-10 giugno (2016)
6. Dotti, N.F., Fratesi, U., Lenzi, C. and Percoco, M. : Local Labour Markets and the Interregional Mobility of Italian University Students, *Spatial Economic Analysis*. Taylor and Francis Journals, 8 (4), 443-68 (2013)
7. Enea, M.: From South to North? Mobility of Southern Italian students at the transition from the first to the second level university degree. Paper presented at the 48th Scientific meeting of the Italian Statistical Society (2016)
8. Giambona, F., Porcu, M., Sulis, I.: Students Mobility: Assessing the Determinants of Attractiveness Across Competing Territorial Areas. *Soc Indic Res*, 133(3) , 1105-1132 (2017)
9. Leckie, G. and Charlton, C.: runmlwin - A Program to Run the MLwiN Multilevel Modelling Software from within Stata. *Journal of Statistical Software*, 52 (11), 1:40 (2013)
10. Pigni, C., and Staffolani, S. Beyond participation: Do the cost and quality of higher education shape the enrolment composition? The case of Italy. *Higher Education*, 71(1), 119142 (2016)
11. Suhonen, T.: Field-of-Study Choice in Higher Education: Does Distance Matter?, *Spatial Economic Analysis*, 9(4), 355-375 (2014)