I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11026

# The 'Gravity' of Quality: Research Quality and Universities' Attractiveness in Italy

Massimiliano Bratti
Stefano Verzillo

SEPTEMBER 2017

DISCUSSION PAPER SERIES

IZA DP No. 11026

# The 'Gravity' of Quality: Research Quality and Universities' Attractiveness in Italy

**Massimiliano Bratti**
*European Commission JRC, University of Milan, IZA and LdA*

**Stefano Verzillo**
*European Commission JRC and University of Milan*

SEPTEMBER 2017



**IZA – Institute of Labor Economics**

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

# ABSTRACT

## The 'Gravity' of Quality: Research Quality and Universities' Attractiveness in Italy*

This paper investigates whether or not research quality is significantly associated with a university's ability to attract students from other provinces in Italy. First university enrolments of students over the period 2003–2011 are regressed on several universities' research quality indicators computed from different bibliometric databases (ISI-Thompson, Scopus-Elsevier and Google Scholar) using fixed effects-gravity models. Our estimates suggest that improving research quality may be an effective way of reducing 'brain drain' from southern Italy.

**JEL Classification:** I23

**Keywords:** brain drain, research quality, student mobility, university, Italy

**Corresponding author:**
Massimiliano Bratti
European Commission Joint Research Centre (JRC)
Via E. Fermi 2749, TP 361
21027 Ispra
Italy
E-mail: massimiliano.bratti@unimi.it

* This version benefited from comments received at the University of Cagliari, the Sixth International Workshop on Applied Economics of Education (IWAEE, Catanzaro), the XXXI National Conference of Labour Economics (Cagliari), the ERSA/RSAI 2015 Barcelona Workshop on Regional and Urban Economics: Spatial Perspectives of Human Capital (Barcelona) and the conference '*Il Sistema universitario: è già 'Buona università?*' (Naples). Comments by Marco Bertoni, Steve Gibbons and Stefano Usai are especially acknowledged. The information and views set out in this paper are those of the authors and do not reflect the official opinion of the European Union. Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use which may be made of the information contained therein. The usual disclaimers apply.

## 1. Introduction

Human capital accumulation is a key factor for economic competitiveness. A highly-educated labour force is indeed more productive (Chevalier et al. 2004, Rosenzweig, 1995), allows effective adoption of new technology (Benhabib & Spiegel, 1994; Comin & Hobijn, 2004), increases innovation and scientific knowledge (Andersson et al. 2009; Waldinger, 2016) and may have positive production spillovers (Moretti, 2004; Rauch, 1993), all factors which positively contribute to a country's economic growth.

The most dynamic economies are successful not only in producing new human capital, but also in attracting highly educated workers from abroad ('brain gain'). The latter can be done either through high skilled workers' immigration or attracting foreign students – typically in tertiary education – who first acquire advanced education and then may find employment in the host country. Besides labour market conditions, cross-country differences in the quality of higher education institutions (HEIs) are another important factor driving international student mobility (A. Abbott & Silles, 2016; Aslangbengui & Montecinos, 1998; Beine et. al., 2014; Gordon & Jallade, 1996).

The capacity to create and attract human capital is also important to explain differences in economic performance across regions of a country (Faggian & McCann, 2009). Economic laggard regions generally have both worse-performing labour markets and a lower quality of HEIs. However, country-level evidence on the role of differential educational quality between regions as a potential determinant of the 'brain drain' is very scant. For the US, some studies have proxied universities' quality with their prestige. A study finds that prestige of undergraduate programmes measured by the Gourman ratings of colleges and universities and the Gross-Grambsch quality ratings of graduate facilities explains only a small proportion of the variation in interstate migration of students (W. R. F. Abbott & Schmid, 1975). Somewhat contrasting are the results of another US study (Baryla & Dotterweich, 2001), which proxies programmes' quality with selectivity of admissions, and finds a positive effect on attraction of non-resident students. A positive association between university quality and regional student mobility is not ubiquitous outside the US either, and is not found, for instance, for the Netherlands by Sá, Florax and Rietveld (2004), who use a composite quality index of educational programmes based on a survey run by the weekly magazine *Elsevier*. Mixed results are also reported for Italy, which is the focus of the current study. Like the results of the papers mentioned above, they may be partly due to the variety of proxies of university quality that scholars have used, which are likely to capture very different aspects of 'quality'. These indicators include, *inter alia*, the proportion of faculty members who received national research funds (Dotti et al., 2013); composite indicators mixing different aspects of university life such as student academic outcomes, teaching quality, research quality and international cooperation (Pigini & Staffolani, 2016); and the ranking obtained in a national research evaluation exercise (Ciriaci, 2014).[1]

This paper seeks to overcome some of the weaknesses of the extant literature and to contribute to the discussion on the role played by university quality in student internal mobility in Italy. The Italian case is interesting from a policy perspective. Southern Italian HEIs are increasingly losing students. The drop in student enrolment numbers throughout the country has been close to 20.4% (66,000 students between 2004 and 2015), with a larger drop in the islands and southern regions (close to 30.2% and 25.5%, respectively) than in the central and northern regions (Viesti, 2016). Part of this

[1] Only the last two quality indicators are found to be positively associated with student choices.

haemorrhage of students from the South may be due to the lower quality of HEIs, which, along with the search for better employment opportunities (Dotti et al., 2013), pushes individuals to move to the North. Our study throws light on this issue by studying student mobility for almost a decade, thoroughly assessing whether or not the lower quality of southern HEIs may be partly responsible for the 'brain drain' that southern Italy is suffering.

Because of the difficulty of building a 'catch-all' indicator of university quality, our analysis only focuses on research quality. Unlike for research, whose quality is generally assessed through bibliometric indicators, there are no obvious quality indicators for teaching. Research quality is proxied using multiple indicators computed from several bibliometric sources. This enables us to check the robustness of our results to variables capturing different aspects of research quality, some of which put more weight on 'productivity', i.e. 'quantity' (e.g. number of publications), and others on the 'quality' (e.g. number of citations) and the impact (e.g. journal impact factor) of research output. Moreover, the use of indicators coming from several sources (ISI – Thomson; Google Scholar; Scopus – Elsevier) allows us to check the sensitivity of the results to employing databases with different levels of coverage of research output. Third, unlike most previous papers, which often focused only on one or a few years, our study uses panel data spanning almost a decade. This enables us to control for universities' time-invariant unobservable characteristics through fixed effects, and to assess the importance of research quality in very different macro-economic conditions. Indeed, our estimation period spans 2003-2011, i.e. including both years before and years after the global financial crisis (GFC). Restricted access to bank credit, less student financial support and worsening of university graduates' labour market prospects following the GFC may have substantially changed the costs and benefits of choosing higher university quality, especially when it is not readily available close to an individual's home.

Last but not least, following the most recent literature (e.g. Chevalier & Jia, 2016; Gibbons et al., 2015) our paper considers student enrolment flows where the units of observations are subject-groups within HEIs. This has some advantages compared with estimating gravity models at either region or province level or even for the same HEI pooling all subject-groups together:[2] averaging research quality indicators may hide substantial heterogeneity existing in the quality of research between HEIs located in the same geographical unit, or between subject-groups within the same HEI.

Our empirical strategy consists in the estimation of a gravity model in which province-level student inflows (i.e. first-time enrolments) by subject-group from the NUTS-3 region (i.e. Italian provinces) of residence (origin) towards an HEI branch (defined at the HEI/destination province level) are regressed on measures of the latter's research quality and a comprehensive set of fixed effects. The results of two specifications are presented. In the first one, the effect of research quality is identified using between-subject-group variation in research quality within the same HEI branch[3] in each year through the inclusion of HEI branch-year fixed effects. To give just one example, this analysis compares the attractiveness of the School of Law of the University of Milan with that of the School of Political, Economic and Social Sciences of the same university, and investigates whether or not differences in student enrolments are partly related to the differential quality of the two schools' research outputs, after accounting for national differences in research quality across subject-groups. In this model, the effect of quality is identified only by universities providing courses in different

[2] For instance, Agasisti and Dal Bianco (2007) and Dotti et al. (2013) use provinces as destinations, while in recent work Cattaneo, Malighetti, Meoli & Paleari (2017) use HEIs as destinations, but do not distinguish student flows by subject-group.

[3] Since each HEI may have branches in more than one province, destinations are defined at the HEI province level.

subject-groups within the same branch. In our estimation sample, for instance, they account for 58%, 71%, 67% and 53% of all university branches in 2011, for northern, central and southern Italy and the islands, respectively.

Results are also reported for a second empirical specification that exploits between-field yearly variation in research quality across HEI branches located in the same province, by including branches' destination province-year fixed effects. In this model, identification stems from those provinces where there are different HEI branches supplying courses in the same subject-group. In both strategies, pull factors associated with differences in destination provinces' attractiveness (e.g. the state of the labour market, amenities, infrastructures) are controlled for through province-year fixed effects. This ensures that research quality can capture a feature of only HEIs, and does not reflect province-level unobservable characteristics.

Because of the presence of a non-negligible fraction of zeros (about 81%) in the dependent variable, following the recommendations of Santos Silva and Tenreyro (2006), gravity models are estimated using Poisson Pseudo Maximum Likelihood (PPML).

The structure of the paper is as follows. Section 2 introduces the conceptual framework, which represents the starting point for the specification of the gravity equation described in Section 3. Section 4 describes our data, and the research quality indicators used in our empirical analysis. The empirical results are commented on in Section 5. Section 6 summarizes the main findings and draws some conclusions.

## 2. Conceptual framework and empirical model

This section introduces a simple random utility model (RUM) that is useful to derive the empirical specification of the gravity equation presented in the next section. RUMs have been already used in the context of migration choices (see Beine et al, 2016 for a review). A version adapted to students' enrolment choices is presented in what follows.

An individual located in province $j$, who has already decided to enrol in higher education at time $t$, has to choose among $k = 1, \ldots, D$ potentially available 'destinations', defined as a combination of HEI's branch ($b$) and field of study ($f$). An HEI ($h$) may have branches in different provinces,[4] and each branch may offer degree courses in several fields of study. The individual's utility can be defined as:

$$U_{ijkt} = w_{jkt} - c_{jkt} + \varepsilon_{ijkt} \qquad (1)$$

where $i$, $j$, $k$ and $t$ are individual, origin-province, destination and time subscripts; $w_{jkt}$ and $c_{jkt}$ are deterministic components of utility observed by the econometrician, reflecting the (pecuniary and non-pecuniary) benefits and costs, respectively, of choosing alternative $k$ for an individual residing in province $j$;[5] and $\varepsilon_{ijkt}$ is an individual-specific stochastic component of utility.

If the random component of utility follows an independently identically Extreme Value Type-1 distribution (also known as Gumbel distribution) and the student chooses the alternative maximizing her utility, then the probability of choosing destination $k$ is:

[4] Provinces are NUTS-3 Italian regions according to the Nomenclature of territorial units for statistics.

[5] Both costs and benefits are assumed not to be individual specific and accordingly are not indexed by $i$.

$$Prob(d_{ijkt} = 1) = \frac{\exp(w_{jkt} - c_{jkt})}{\sum_{l=1}^{D} \exp(w_{jlt} - c_{jlt})} \quad (2)$$

where $d_{ijkt}$ is an indicator variable which takes on the value 1 if the destination *k* is chosen at time *t* by individual *i* residing in province *j* and 0 otherwise.

Aggregating the individual choices at the origin-province level, the number of students coming from province *j* who enrol in tertiary education in *k* at time *t* can be written as:[6]

$$S_{jkt} = Prob(d_{ijkt} = 1)P_{jt} = \phi_{jkt} \frac{y_{jkt}}{\Omega_{jt}} P_{jt} \quad (3)$$

where $\phi_{jkt} = \exp(-c_{jkt})$, $y_{kjt} = \exp(w_{jkt})$, $\Omega_{jt} = \sum_{l=1}^{D} \phi_{jlt} y_{jlt}$ and $P_{jt}$ is the province *j* population 'at risk' of becoming migrant students at time *t* (composed of students who have decided to continue in higher education). Equation (3) suggests that the flow of students from *j* to *k* depends positively on *k*'s accessibility with respect to *j* ($\phi_{jkt}$), on its attractiveness ($y_{jkt}$) and on a multilateral resistance term ($\Omega_{jt}$) that depends on the accessibility and attractiveness of the competing destinations. This last term captures, for instance, that a decrease in the attractiveness of destinations competing with destination *k* increases *k*'s relative attractiveness.

Equation (3) can also be rewritten as:

$$S_{jkt} = \exp(\ln(P_{jt}) + w_{jkt} - c_{jkt} - \ln(\Omega_{jt}) \quad (4)$$

(subscripts: *j*=origin province; *k*=destination; *t*=time)

with $S_{jkt} \geq 0$, whose stochastic counterpart is:

[6] By the law of large numbers, the probability that a student residing in province *j* chooses to enrol at university in destination *k* approximately coincides with the fraction of students of province *j* enrolling in destination *k* (see Ortega and Peri 2012).

$$S_{jkt} = \exp(\ln(P_{jt}) + w_{jkt} - c_{jkt} - \ln(\Omega_{jt}))h_{jkt} \quad (5)$$

(subscripts: *j*=origin province; *k*=destination; *t*=time)

where $h_{jkt}$ is an error term with $\mathrm{E}(h_{jkt}) = 1$.

Model (5) can be estimated using the PPML estimator (Santos Silva & Tenreyro, 2006).[7]

### 3. The gravity equation

We operationalize equation (5) as follows. The dyadic costs of choosing destination *k* for students coming from *j* are:

$$c_{jkt} =$$
$$a_0 + a_1 lnDIST_{jb} + a_2 lnDIST_{jb} * SEA_b + a_3 SEA_b +$$
$$a_4 CONTIG_{jb} + a_5 SAMEREG_{jb} + a_6 SAMEPROV_{jb} + a_7 C_{bt} + a_8 C_{jt} + a_9 C_{ft} \quad (6)$$

(subscripts: *j*=origin province; *k*=destination; *f*=subject-group; *t*=time)

where $lnDIST_{jb}$ is the geodesic distance between the centroids of the origin province and of the province where the university branch is located in logarithm (this variable has the same value for all destinations, i.e. branch-subject-group combinations, within a branch); $lnDIST_{jb} * SEA_b$ is an interaction term between distance and an indicator $SEA_b$ for either the province of origin or that of the destination branch being located in an island (Sicily or Sardinia), which captures the higher travel costs related to sea crossing; $CONTIG_{jb}$ is a contiguity indicator between the two provinces;

[7] PPML estimation, unlike log-linear models, is particularly suitable when there are many zeros in the dependent variable (about 81% in our estimation sample). This estimation method has become very popular for gravity models of trade and immigration.

$SAMEREG_{jb}$ is a dichotomous indicator for origin and destination provinces being both located in the same region; $SAMEPROV_{jb}$ is a dichotomous indicator for province of origin and university branch of destination being both located in the same province.[8] The variables measuring provinces' contiguity and provinces in the same region capture potential non-linear effects of distance. $C_{bt}$, $C_{jt}$ and $C_{ft}$ capture time-varying cost components with variation at the levels of the HEI's branch, students' province of origin and subject-group, respectively. Variables such as the average level of unemployment, unskilled wages and living costs in the origin province are captured by $C_{jt}$, while the average cost of living in the province where *b* is located is captured by $C_{bt}$. Finally, $C_{ft}$ captures factors such as differences in average enrolment fees across subject-groups at the national level, and higher selectivity in student admissions for specific subject-groups.[9]

Students' benefits of choosing destination *k* are modelled as:

$$w_{bkt} = b_0 + b_1 lnQ_{kt} + b_2 B_{ft} + b_3\ lnSTU\_TEACH_{hft} \qquad (7)$$

(subscripts: *j*=origin province; *k*=destination; *h*=HEI; *f*=subject-group; *t*=time)

which depends on the logarithm of research 'quality' of destination $k$ at time $t$, i.e. $lnQ_{kt}$; the average benefits from enrolling in field *f* at the national level ($B_{ft}$); and the average benefits of (log) teaching quality, proxied by the student-teacher ratio ( $lnSTU\_TEACH_{hft}$), which is included at the HEI-subject-group level.[10] The rationale for including research quality in the expected benefits of graduating from destination $k$

---

8 Coherently with the RUM of Section 2, in which a student has to choose among all available destinations, enrolments of students coming from the province where the university branch is located are also included in the analysis (0.91% of the between-provinces flows, which accounts, however, for 53.9% of total enrolled students).

9 Unfortunately, in our data, average fees cannot be computed by either field of study or branch, and they vary only at the HEI level.

10 Since it is not possible to link teaching staff to the branches in which they actually taught (in pure teaching branches), the student-teacher ratio is measured at the HEI-subject-group level.

$(w_{kt})$ is the existing evidence that it positively affects university graduates' wages and employability (Ciriaci & Muscio, 2014; Di Pietro & Cutillo, 2006).

Thus equation (5) can be rewritten as:

$$S_{jkt} = \exp(\ln(P_{jt}) + b_0 + b_1 lnQ_{kt} + b_2 B_{ft} + b_3\ lnSTU\_TEACH_{hft} - a_0 - a_1 lnDIST_{jb} - a_2 lnDIST_{jb} * SEA_b - a_3 SEA_b - a_4 CONTIG_{jb} - a_5 SAMEREG_{jb} - a_6 SAMEPROV_{jb} - a_7 C_{bt} - a_8 C_{jt} - a_9 C_{ft} - \ln(\Omega_{jt}))\ h_{jkt} \quad (8)$$

(subscripts: *j*=origin province; *k*=destination; h=HEI; *f*=subject-group; *t*=time).

As measures of the variables $B_{ft}$, $C_{bt}$, $C_{jt}$ and $C_{ft}$ are not available, they are proxied with fixed effects ($D_{.t}$). So, after noting that $\ln(P_{jt})$ and $\ln(\Omega_{jt})$ exhibit variation at the origin province by time level, equation (8) becomes:

$$S_{jkt} = \exp(\gamma_0 + b_1 lnQ_{kt} + b_3\ lnSTU\_TEACH_{hft} - a_1 lnDIST_{jb} - a_2 lnDIST_{jb} * SEA_b - a_3 SEA_b - a_4 CONTIG_{jb} - a_5 SAMEREG_{jb} - a_6 SAMEPROV_{jb} + D_{ft} + D_{bt} + D_{jt})\ h_{jkt} \quad (9)$$

(subscripts: *j*=origin province; *k*=destination; *h*=HEI; *f*=subject-group; *t*=time)

where $\gamma_0 = b_0 - a_0$.

Our coefficient of interest is $b_1$ measuring the 'return to university research quality.'[11]

Several research quality indicators are considered, some of which measure more

[11] If $lnQ_{kt}$ and $lnSTU\_TEACH_{jb}$ are also included in the cost of university of enrolment, then $b_1$ and $b_3$ would capture the net return to university's research quality and teaching quality, respectively.

quantitative and others more qualitative dimensions of research output. By including time-variant university-branches fixed effects ($D_{bt}$), only between-field variation in research quality within the same university branch is exploited in each year. The fixed effects $D_{ft}$ capture *inter alia* average differences in productivity and impact of research between subject-groups.[12]

Since equation (5) is likely to mainly identify the effect of research quality in HEIs that offer courses in multiple fields within the same branch, an alternative specification exploiting within-field variation across HEI's branches located in the same province *p* is also estimated:

$$S_{jkt} = \exp(\gamma_0 + b_1 lnQ_{kt} + b_3\, lnSTU\_TEACH_{hft} - a_1 lnDIST_{jb} - a_2 lnDIST_{jb} * SEA_b - a_3 SEA_b - a_4 CONTIG_{jb} - a_5 SAMEREG_{jb} - a_6 SAMEPROV_{jb} + D_{ft} + D_{pt} + D_{jt})\, h_{jkt} \tag{10}$$

(subscripts: *j*=origin province; *p*=destination province; *k*=destination; *h*=HEI; *f*=subject-group; *t*=time)

where the destination branch-year fixed effects in equation (9) have been replaced with destination province-year fixed effects.

In the PPML estimates, coefficients of variables in logarithm can be roughly interpreted as elasticities and coefficients on variables in levels as percentage changes in the dependent variable.

[12] In a recent paper, Cattaneo et al. (2017) investigate the effect of spatial competition among universities. In our models, spatial competition factors are captured by the multilateral resistance term, proxied by destination-province fixed effects. Moreover, competitors' proximity indexes similar to those included by Cattaneo et al. are subsumed in our models in branch-year fixed effects.

## 4. Data

Six main data sources are used in the empirical analysis. The first one is the administrative archive of the Italian Ministry of Education, University and Research (MIUR), gathering information (scientific field, university affiliation and academic rank) on the whole set of individuals holding an academic position in Italian universities (Assistant Professors, Associate Professors and Full Professors) for the whole estimation period. A second source provides information on the flows of first-time enrolments by students' province of residence before enrolment for each degree course run by Italian universities (the dependent variable in the gravity model) and some control variables used in the regressions, such as the teacher-student ratio. The analysis in this paper is limited to first enrolments in first-level and single-cycle degrees. Indeed, a student's province of residence is likely be a better indicator of her provenience while she is attending secondary education (i.e. when she still lives with her parents) than after completing a first degree and when enrolling in postgraduate education, as many students in Italy do not change their legal residence during university studies. Degree-level data are aggregated in university branch-subject-group-year cells, which represent the statistical units of analysis. Three fee-based and open-access research repositories, namely the ISI Web of Science (WoS), Scopus (Elsevier) and Google Scholar (GS) archives, are used to build a set of research quality indicators related to both individual productivity (e.g. number of papers published) and scientific impact (e.g. number of citations). Finally, a sixth data source used in this study is Thomson Journal Citation Reports, which collects impact factors of scientific journals.

Table 1 provides the list of the variables used in our analysis, with their names, descriptions and data sources, and Table 2 presents descriptive statistics of all the variables included in the final model.

[Table 1 near here.]

[Table 2 near here.]

### *4.1 Research quality indicators*

This section reports the definition of the research quality indicators used in the empirical analysis, while a detailed description of the procedure followed for their construction is given in Appendix C.

The analysis of universities' attractiveness proposed in this paper is based on the estimation of PPML gravity models using observations corresponding to the cells defined by students' province of origin (proxied by student residence at the time of enrolment) and the province, field of study and HEI of the degree course in which they enrolled, where the pair province-HEI defines a university branch. Consequently, the model estimation requires computing for each university branch and scientific field some indicators of research quality.

Unfortunately, information is not available to link the flows of students attracted by each branch-subject-group (i.e. destination) with the exact pool of researchers and professors providing them with teaching activities. Hence, student flows were assigned the research quality indicators of all professors and researchers belonging to the same HEI and the same 'broad scientific subject-groups' of the enrolled course. 'Broad scientific subject-groups' are defined as the best lexicographical match between the subject-group classification of university courses provided by MIUR for teaching purposes and the scientific research areas according to which the academic research staff are hired ('scientific-disciplinary areas' as defined by the National University Council, CUN).[13] The correspondence table between teaching-related subject-groups (MIUR) and scientific-related subject-groups is provided in Table A1 in Appendix A.

Taking advantage of the abovementioned bibliometric sources, the following seven research quality indicators were computed:

1) *Average number of ISI publications*$_{fht}$: the weighted average of the total number of publications recorded in WoS for each researcher in field *f* affiliated with institution *h* at time *t*;
2) *Average number of ISI citations*$_{fht}$: the weighted average of the total number of citations in WoS for each researcher in field *f* affiliated with institution *h* at time *t*;
3) *Average impact factor of ISI publications*$_{fht}$: the weighted average of the impact factor (Journal of Citation Reports) of the WoS publications of each researcher in field *f* affiliated with institution *h* at time *t*;
4) *Average number of GS publications*$_{fht}$: the weighted average of the total number of publications on Google Scholar by each researcher in field *f* affiliated with institution *h* at time *t*;
5) *Average number of citations of GS publications*$_{fht}$: the weighted average of the cumulative numbers of citations on Google Scholar for each researcher in field *f* affiliated with institution *h* at time *t*;
6) *Average number of Scopus publications*$_{fht}$: the weighted average of the total number of publications on Scopus by each researcher in field *f* affiliated with institution *h* at time *t*;

[13] A reader who is not familiar with Italian academe might not be aware of the fact that each academic is allocated to only one out of 370 scientific sectors (*settori scientifico disciplinari*), sectors that are mostly relevant for career progression because the hiring and promotion procedures are carried out within these sectors.

7) *Average number of citations of Scopus publications*$_{fht}$: the weighted average of all the individual cumulative number of citations on Scopus for each researcher in field *f* affiliated with institution *h* at time *t*.

When 'broad scientific subject-groups' encompasses more than one CUN scientific group, weighted averages were computed according to the number of first-level enrolled students belonging to each scientific area (CUN). Moreover, pure teaching branches, i.e. branches that do not have formally staff linked to them, were attributed the research quality of their headquarters.

Table B1 in Appendix B reports the pairwise correlations between the different research quality indicators. Although indicators from the same source are highly correlated, the same is not true of indicators from different sources, suggesting that they have some independent variation.

## 5. Empirical results

### *5.1 Time-variant university 'branch' (i.e. HEI-province) fixed effects*

This section reports the results of the model including time-variant university 'branch' fixed effects, i.e. the gravity equation (9). In this model, identification comes from differences in research quality between subject-groups of degree courses provided by the same university branch in a given year.

The estimates in all columns of Table 3 use the same control variables, but different indicators of research quality. Our models explain around 83% of the total variance in student inflows. Results show the expected signs for the main control variables included. A negative and statistically significant relationship emerges between the student inflows and geodesic distance between the student's province of residence and the province of the degree course in which the student is enrolled, with an elasticity very close to minus unity. The elasticity of student inflows with respect to distance decreases by about -0.35 for travels that involve sea crossing.

As expected, province contiguity is positively associated with student inflows, with contiguous provinces enjoying a premium of about 1% in student inflows (1.11%-1.12% depending on the research quality indicator used). This confirms the students' attitude of moving from their province of origin to a destination in a nearby province rather than somewhere else in Italy. Inflows of students within provinces in the same region are more likely to occur (1.16% more) than flows involving provinces located in different regions. In all the estimated models of Table 3, the coefficients of these control variables are highly stable in magnitude. The negative coefficient on within-province inflows suggests that branches, on average, attract more students from outside the province than within the province where they are located (on average around 5.3% more). To interpret this effect, however, it is worth keeping two points in mind. (i) These estimates are comparing between-subject-group variation within the same branch, that different subject-groups may compete for the same local students. Local students may have pondered their choice of subject less than those coming from other provinces. (ii) This negative coefficient may partly capture the effect of distance, which is set to zero for within-province enrolments.

Teaching quality is proxied by the (log of) student-teacher ratio, as is customary in this literature (Agasisti & Dal Bianco, 2007; Ciriaci, 2014). The estimates show a statistically significant elasticity of 0.19. The result of a *positive* effect of the student-teacher ratio on student inflows can be explained by the mechanical positive correlation between degree courses in high demand (i.e. 'crowded' courses) and student inflows.

Coming now to our variables of interest, i.e. the research quality indicators, Table 3 demonstrates the high relevance of research quality in explaining student flows.

[Table 3 near here.]

All research quality indicators are positively associated with university branches' attractiveness, and statistically significant at least at the 1% level. Student inflows show a 0.026 elasticity to *research productivity*, proxied by the number of WoS publications (column 1). The elasticity is somewhat smaller when *research influence* is considered, using the average number of WoS citations and the average impact factor, with elasticities of 0.013 and 0.017 in columns (2) and (3), respectively. Interestingly, the effects appear to be larger when GS's research quality indicators are used, with elasticity of student flows to scientific productivity and impact of 0.028 (column 4) and 0.034 (column 5), respectively. This may be partly explained by the higher accessibility to students of the scientific production surveyed by GS, which has a much better coverage of articles and books written in Italian, and the higher inclusivity of GS for subject-groups in the arts and humanities. Last, the elasticity of student enrolments is largest (0.059) when the average number of Scopus publications is considered (column 6). The elasticity of student inflows to the number of Scopus citations lies instead between those estimated with WoS and GS (0.026).

The period covered by our empirical investigation (2003-2011) is quite heterogeneous. In particular, the onset of the GFC in 2007-2008 may have changed the relevance of university research quality in student enrolment choices. The direction of the change is unknown a priori. On the one hand, owing to its public debt burden, Italy saw an overall contraction of the higher education system in terms of both financial and human resources (Viesti, 2016) after the crisis. This might have entailed a higher tendency of students to enrol in the local university irrespective of university quality. On the other hand, students and their families might have put more weight on university quality in response to tougher competition in the university graduates' labour market determined by a depressed demand.

To investigate these hypotheses, Tables D1 and D2 (in Appendix D) split the estimation sample into the pre- and post-crisis periods. Except for the lack of significance of the average number of ISI citations and the average impact factor after the GFC – which may be driven by the fact that they need a longer period to consolidate – all other research quality indicators turn out to be positively associated with student inflows in both periods. Interestingly, the research quality indicators using GS and Scopus exhibit larger elasticities in the post-crisis period (e.g. 0.046 versus 0.086 for Scopus number of publications, and 0.019 versus 0.042 for Scopus citations) supporting a bigger role for quality after the GFC.

### *5.2 Time-variant destination-province fixed effects*

Table 4 shows the estimates of the gravity equation exploiting the within-subject-group variation between HEI branches located in the same province, i.e. equation (10).

[Table 4 near here.]

These estimates are reported for the sake of completeness, although compared with those in the previous section they are less robust to the presence of HEI-branches unobserved characteristics potentially correlated with attractiveness. They are nonetheless informative, as they show whether students tend to choose among HEI-branches in the same subject-group and province according to research quality.

The results are consistent with those in the previous section, although the elasticities are all larger in magnitude. The estimated elasticities are 0.131, 0.085 and 0.106 for WoS number of publications, number of citations and impact factor, respectively; 0.16 and 0.146 for GS number of publications and citations, respectively; and 0.145 and 0.083 for Scopus number of publications and citations, respectively. As we expected, the larger point estimates than those in Table 3 may partly reflect HEI-specific factors that are omitted from the regression. Interestingly, the coefficients of the control variables are very close to those in Table 3.

Tables D3 and D4 in Appendix D report the estimates for the pre- and post-crisis period. Also in this case, like in Tables D1 and D2, statistically significant positive effects are found before and after the GFC, with larger estimated elasticities for the most recent period.

### *5.3 Analysis by geographic macro-areas*

In Italy there is an on-going lively debate on the research quality gap between universities in the northern and southern regions of the country and how this may exacerbate the long-standing phenomenon of brain drain from the South. Indeed, those students who go to northern regions to study seldom return to work in their regions of origin because of poor employment opportunities.

Geographical heterogeneity is investigated in Table D5 in Appendix D, which shows the estimates of gravity equations (9) and (10) for four different geographical macro-areas: North, Centre, South and Islands. Starting with the model with time-variant university branches fixed effects (FEs), statistically significant positive elasticities are estimated for all three WoS research indicators (0.038, 0.035 and 0.038 for number of publications, citations and impact factor, respectively) and for two Scopus indicators (0.056 and 0.033 for number of publications and citation, respectively). These are the most selective research quality indicators, since articles and books have to pass minimum standard criteria to be included in the WoS and Scopus databases. For the South, the only notable differences are the lack of significance of the WoS average number of citations, and the statistical significance of the GS research quality indicators, which have large elasticities (0.101 and 0.074, for GS number of publications and citations, respectively). The lack of significance of research quality indicators for the Centre is curious; there, only the Journal of Citation Reports' impact factor is statistically significant, but surprisingly with a negative sign. Somewhat more 'mixed' is the picture emerging for Islands (Sicily and Sardinia), where statistically significant positive effects are estimated for GS publications and citations, while the average number of WoS citations and impact factor are *negatively* associated with student inflows. On the grounds that research-active subject-groups also have higher teaching standards (i.e. are more difficult), this may reflect, depending on HEI choice, a tendency in the Italian islands to enrol in subject-groups that maximize a student's likelihood to complete the degree course. All in all, columns (1)-(4) of Table D5 suggest an important role for research quality in student flows in both northern and southern regions, with less selective GS indicators playing a bigger role in the latter presumably owing to the higher frequency of the publication outlets covered by GS.

Columns (5)-(8) of Table D5 show the estimates of the model with time-variant province-year FEs. The results are qualitatively similar to those in the previous four columns, but with larger estimated elasticities. The only two differences worth noting are the larger elasticities estimated for the Centre, which are now all statistically significant at the 1% level, and the disappearance of the negative effects for the Islands. Like the first result, this could be partly explained by the source of identification of this model, which exploits between-branch differences within subject-groups in the same province, as the Centre is the region where university branches are relatively most geographically dispersed and branches of different HEIs often coexist within the same province (Bratti, Checchi & de Blasio 2008).

## 6. Concluding remarks

This study employs gravity models to investigate the impact of research quality on students' internal mobility using Italian data over the period 2003-2011. Using panel data and framing the analysis at the subject-group level, it allows, for the first time for Italy, the exploitation of within-subject-group differences in research quality across university branches, controlling, *inter alia*, for potential unobservable variables at the HEI-branch level (e.g. alma mater's reputation). Using a large set of fixed effects, the gravity model's estimates provide evidence that stronger research performance is associated with larger inflows of university students. According to our preferred estimates in Table 3 the elasticity of student inflows to research quality varies between 0.013 and 0.059 depending on the research quality indicator used. Our analysis suggests that research quality played a bigger role after the GFC and that more selective quality indicators (WoS, Scopus) are associated with larger student inflows especially in the North, while more comprehensive indicators (GS) predict student inflows in the South.

The results point to the importance of improving research as a possible means through which southern Italy's universities may reduce the migration of students towards northern Italy, and the brain drain associated with it. Using, for instance, the estimated elasticity of student inflows to the number of GS citations simple back-of-the-envelope computations indicate that closing the average gap in research quality between the North and the South (Islands) during the period – e.g., corresponding to a 16.8% (15.5%) increase in GS citations – would increase the annual student inflows in the HEIs located in southern Italy (islands) from the 'average' Italian province by 1.2% (0.8%).[14]

Our results are relevant for the recent debate on the 'Fund for the financing of excellent university departments' ('*Fondo per il finanziamento dei dipartimenti universitari di eccellenza'*) introduced by Law No 232/2016 (also known as 'Stability Law 2016'), providing 1,350,000 euros of additional funding per year, for five years, to the 180 top-ranked departments in the most recent Italian Research Evaluation Exercise (*Valutazione della Qualità della Ricerca*, VQR 2011-2014).[15] In particular, in a situation of declining resources devoted to the university system (Viesti, 2016), Performance-Based Funding Systems based on research performance, although may allow for an effective allocation of financial resources when the objective is that of increasing research productivity and quality, may also increase the gap in research performance between north and south Italy's HEIs, further reducing the attractiveness of the latter to university students.

[14] These figures are obtained by multiplying the percentage change at the sample mean of the geographic area for the whole estimation period in the research quality indicator (GS citations) necessary to close the gap with the North by the estimated elasticities of student inflows to research quality for southern Italy and the islands shown in column (3) and (4) of Table D5, respectively.

[15] More precisely, success in the competition depends both on performance in the VQR 2011-2014 and on the rating of a 'department development plan' submitted to the Ministry of Education, University and Research. The base provision of 1,350,000 euros is reduced or increased according to the quintiles of ranking of the winning departments.

The analysis in this paper shares some of the limitations of the existing literature. First, like previous studies (e.g. Ciriaci, 2014) our paper does not allow the dissection of the micro-level mechanisms through which students and their families become informed about research quality. Although some scientific discoveries are covered by national or local media (television, newspapers, social media), this involves just a small part of the whole research activity. For the diffusion of universities' research outputs to the wider public, an important role is played by research-dissemination activities and the so called 'third mission', which received an important recognition in the most recent VQR. Information on the research quality of specific HEIs, or more broadly on HEI schools' research reputation, can also be gathered by prospective students through contacts with their older peers coming from the same province who are already enrolled in tertiary education, through a mechanism that resembles that of 'immigrant enclaves'.

Second, like all studies investigating student mobility, our empirical models lack robust indicators of teaching quality. Although a proxy of the latter is included (the student-teacher ratio), our estimates may still suffer from an omitted variables bias. The current analysis remains in our opinion informative. Indeed, if there is a positive correlation between teaching and research activities (i.e. if they are complementary activities, or in other words if more research makes you a 'better' teacher), the estimated elasticities provided in this paper could be interpreted as those of overall 'university quality' encompassing both teaching and research quality. In this case, our results would suggest that it is difficult to implement a dual higher education system in which HEIs are divided into 'teaching' and 'research' universities, and for an HEI not active in research to attain teaching excellence. On the contrary, if teaching and research are substitutes (i.e. negatively correlated, i.e. investing more time in teaching makes you a 'worse' researcher), our estimates would only provide lower bound estimates of the true effect of research quality on student mobility. It is difficult to determine the bias *a priori*, as it would mainly depend on the nature of substitutes or complements of teaching and research. However, the existence of an individual's time constraint, the time-intensive nature of both activities, the strong emphasis given to research productivity for career promotion by recent reforms – such as the introduction of VQR and the National Scientific Habilitation (ASN[16]) of professors – and by the market (e.g. through income from consultancy or research funding) make it more likely that the two activities are substitutes for each other, i.e. better research staff would have higher opportunity costs of investing more time in teaching quality. Recent research exploiting quasi-experimental evidence from Italy (De Philippis, 2015) seems indeed to confirm this view.

[16] The ASN is a centralized procedure established in 2010 (by Law No 240/2010). It is a selective procedure whose outcome is a list of habilitated researchers who can participate in the local selection procedures for associate or full professors managed by Italian universities.

**References**


Abbott, W. R. F., & Schmid, C. F. (1975). University prestige and first-time undergraduate migration in the United States. *Sociology of Education, 48*(2), 168–185.

Abbott, A., & Silles, M. (2016). Determinants of international student migration. *World Economy, 39*(5), 1467–19701.

Agasisti, T., & Dal Bianco, A. (2007). Determinants of college student migration in Italy: Empirical evidence from a gravity approach. http://dx.doi.org/10.2139/ssrn.1063481

Andersson, R., Quigley, J. M., & Wilhelmsson, M. (2009). Urbanization, productivity, and innovation: Evidence from investment in higher education. *Journal of Urban Economics, 66*(1), 2–15.

Aslangbengui, N., & Montecinos, V. (1998). Foreign students in US doctoral programs. *Journal of Economic Perspectives, 12*(3), 171–182.

Baryla, E., & Dotterweich D. (2001). Student migration: Do significant factors vary by region? *Education Economics, 9*(3), 269–280.

Beine, M., Noël, R., & Ragot, L. (2014). Determinants of the international mobility of students. *Economics of Education Review, 41*(C), 40–54.

Beine, M., Bertoli, S., & Moraga, J. F.-H., (2016). A practitioners' guide to gravity models of international migration. *World Economy, 39*(4), 1467–19701.

Benhabib, J., & Spiegel, M. M. (1994). The role of human capital in economic development evidence from aggregate cross-country data. *Journal of Monetary Economics, 34*(2), 143–173.

Bratti, M., Checchi, D., & de Blasio, G. (2008). Does the expansion of higher education increase the equality of educational opportunities? Evidence from Italy. *Labour, 22*(S1), 53–88.

Cattaneo, M., Malighetti, P., Meoli, M., & Paleari, S. (2017). University spatial competition for students: the Italian case. *Regional Studies, 51*(5), 750–764.

Chevalier, A., Harmon, C., Walker, I., & Zhu, Y. (2004). Does education raise productivity, or just reflect it? *Economic Journal, 114*(499), 499–517.

Chevalier, A., & Jia, X. (2016). Subject-specific league tables and students' application decisions. *Manchester School, 84*(5), 600–620.

Ciriaci, D. (2014). Does university quality influence the interregional mobility of students and graduates? The case of Italy. *Regional Studies, 48*(10), 1592–1608.

Ciriaci, D., & Muscio, A. (2014). University choice, research quality and graduates' employability: Evidence from Italian national survey data. *European Educational Research Journal, 13*(2), 199–219.

Comin, D., & Hobijn, B. (2004). Cross-country technology adoption: Making the theories face the facts. *Journal of Monetary Economics, 51*(1), 39–83.

Di Pietro, G., & Cutillo, A. (2006). University quality and labour market outcomes in Italy. *Labour, 20*(1), 37–62.

De Philippis, M. (2015). *Multitask agents and incentives: The case of teaching and research for university professors* (CEP Discussion Papers dp1386). Centre for Economic Performance, LSE, London, UK.

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

Faggian, A., & McCann, P. (2009). Human capital, graduate migration and innovation in British regions. *Cambridge Journal of Economics, 33*(2), 317–333.

Gibbons, S., Neumayer, E., & Perkins, R. (2015). Student satisfaction, league tables and university applications: Evidence from Britain. *Economics of Education Review, 48*(C), 148–164.

Gordon, J., & Jallade, J. (1996). “Spontaneous” student mobility in the European Union: A statistical survey. *European Journal of Education, 31*(2), 133–151.

Moretti, E. (2004). Workers’ education, spillovers, and productivity: Evidence from plant-level production functions. *American Economic Review, 94*(3), 656–690.

Ortega, F., & Peri, G. (2012). *The role of income and immigration policies in attracting international migrants* (No 6655, IZA Discussion Papers). Institute for the Study of Labor (IZA), Bonn, Germany.

Pigini, C., & Staffolani, S. (2016). Beyond participation: Do the cost and quality of higher education shape the enrolment composition? The case of Italy. *Higher Education, 71*(1), 119–142.

Rauch, J. E. (1993). Productivity gains from geographic concentration of human capital: Evidence from the cities. *Journal of Urban Economics, 34*(3), 380–400.

Rosenzweig, M. R. (1995). Why are there returns to schooling? *American Economic Review, 85*(2), 153–158.

Sá, C., Florax, R., & Rietveld, P. (2004). Determinants of the regional demand for higher education in The Netherlands: A gravity model approach. *Regional Studies, 38*(4), 375–392.

Santos Silva, J. M. C., & Tenreyro, S. (2006). The log of gravity. *Review of Economics and Statistics, 88*(4), 641–658.

Viesti, G. (Ed.). (2016). *Università in declino: Un Rapporto sugli atenei italiani da Nord a Sud*. Fondazione Res. Donzelli Editore. Rome.

Waldinger, F. (2016). Bombs, brains, and science: The role of human and physical capital for the creation of scientific knowledge. *Review of Economics and Statistics, 98*(5), 811–831.

## Supplemental Online Material (Appendices)

### A. Mapping of teaching-related and research-related subject-groups

[Table A1 near here.]

### B. Correlation between research quality indicators

[Table B1 near here.]

### C. Construction of research quality indicators

#### *C.1. ISI Web of Science and Scopus (Elsevier)*

The data collected from the web versions of WoS and Scopus have information on each academic article published by at least one author affiliated with an Italian HEI. The datasets listed more than 1 million publications each, with details on titles, authors' surnames and first initials, affiliations, journal names and their ISSN (international standard serial number) codes. We have also matched the ISI dataset to the Journal Citation Reports 2012 using the ISSN journal code to match each publication with standard bibliometric measures at the journal level, such as the impact factor (measured over the period 2008-2012) and the research discipline as it is defined by Thomson Reuters. The resulting dataset required a considerable amount of cleaning work to detect duplicates, incomplete records and homonyms. A complete and detailed overview of the data cleansing process is described by Checchi et al. (2014).

In this appendix, we briefly summarize the procedure already proposed and used by Checchi et al. (2014) to correct homonyms for scientific papers whose authors have the same surname, first initial of the name and affiliation. First, scientific sectors of Italian academia were aggregated into 29 groups using the alphabetic part of their institutional codes (e.g. if the scientific sector codes for Statistics, Experimental Statistics, Economic Statistics and Demography are, respectively, SECS-S/01, SECS-

S/02, SECS-S/03 and SECS-S/04 then their 'alphabetic group' is defined as SECS-S). Subsequently, the characterization of journals in 260 'sub-disciplines' provided by the Journal Citation Report 2012 was used to calculate both the probability that a researcher in each of the 29 research areas would publish a paper in a journal categorized in each of the 260 sub-disciplines, and the probability that a paper authored by a researcher belonging to each research area would be published in each of the possible JCR sub-disciplines. Finally, an article authored by a given Italian researcher was assigned to that specific researcher (according to the surname, first initial and affiliation) if and only if either these probabilities exceed a defined threshold given by a proportion of the Herfindal index. Thresholds are specific for each combination of research area and sub-discipline (see Checchi et al., 2014, for more details). In case of $n$ (with $n \geq 2$) homonymous researchers in the same disciplinary area, each paper with the same combination of surname, first initial and affiliation is attributed arbitrarily among them (it is assigned randomly with probability $\frac{1}{n}$ to one of them).

### *C.2 Google Scholar*

Google indexes journal articles, books chapters, conference proceedings, working papers and U.S. patents in its Scholar database. It represents a unique source of free bibliometric information outside the Thomson-Reuters (ISI) – Elsevier (Scopus) fee-based private market. Generally, GS presents a broader picture of researchers' publication records than Scopus or ISI WoS because it covers a wider range of publication outlets (based nationally and internationally) and, consequently, a higher number of citations. Several studies debate the pros and cons of GS in great detail. In this paper, the correlations of our GS metrics and research quality at the institution level are assumed to be reasonably high, like the finding of Smith (2008) that there was a correlation of 0.94 between citation counts from GS and the research output from universities under the New Zealand PBRF (Performance-Based Research Funding) research assessment exercise. To this purpose, a complex algorithm was developed to download and parse GS records of Italian academics (as they are listed in the administrative database provided by MIUR), obtaining a dataset of around 4 million GS bibliometric records. An ad hoc procedure to disentangle possible misattributions of scientific publications due to homonyms was also studied for GS (Ferrara, Montanelli & Verzillo, 2017).

In brief, each author's publications are firstly divided into homogeneous groups on the basis of the frequency of their co-authorship relations. The purpose of this preliminary data analysis is to identify sub-sets of publications belonging to potentially different authors, even if they have the same surname. Subsequently, these groups of papers are revised in a second stage, with two purposes: on the one hand to remove from any group the publications that are not homogeneous (although characterized by co-author links) and on the other hand to merge potentially homogeneous groups of publications although they are initially divided into distinct groups. To this end, a measure of similarity between publications is calculated using natural language-processing and text-analysis tools widely used in fields such as information retrieval and text mining. This measure is mainly based on the overall number of authors, the publication year, the publication venue and the keywords identified in the publications' titles. The result of this second step is a sub-set of groups that contain the name of each author under examination. In order to keep only the groups of publications associated with the real identity of the single author, an automatic comparison of each group of publications with some keywords extracted from the publications enclosed in the curricula vitae of all the individuals who applied for the Italian ASN in 2012 and belong to the same scientific sector as the author in question has been developed by the authors. This procedure allows to obtain a 1.5 million database of disambiguated GS publications used to compute research quality indicators at the university level.

**D. Additional Results**

**[Tables D1, D2, D3, D4 and D5 near here.]**

## References

Checchi, D., De Fraja, G., & Verzillo, S. (2014). *Publish or perish: An analysis of the academic job market in Italy* (IZA Discussion Papers 8345). Institute for the Study of Labor (IZA), Bonn.

Ferrara, A., Montanelli, S., & Verzillo, S. (2017). Google Scholar as a citation database for non-bibliometric areas: The EVA project results. In A. Bonnaccorsi (Ed.), *The evaluation of research in social sciences and humanities*. Springer (*forthcoming*).

Smith, A. G. (2008). Benchmarking Google Scholar with the New Zealand PBRF research assessment exercise. *Scientometrics, 74*(2), 309–316.

## Table 1: Description and source of explanatory variables

| Variables | Description | Data source |
|---|---|---|
| *Student flow* | Flow of students from a province towards a destination (i.e. HEI's branch-subject-group) for all first-level and single-cycle degrees | MIUR |
| *Distance (km)* | Geodesic distance (shortest route between the two centroids) of each pair of provinces | Our computation on the coordinates of provinces centroids provided by ISTAT |
| *Sea* | Dichotomous indicator for either the origin or the destination province being an island | Our computation on the dataset |
| *Sea × Distance* | Interaction effect between Distance and Sea | Our computation on the dataset |
| *Province contiguity* | Dummy variable taking the value 1 for each pair of provinces sharing at least 1 km of border | ISTAT (2012) |
| *Flow within same region* | Dichotomous indicator for student inflow within the same region where the university branch is located | Our computation on the dataset |
| *Within province* | Dichotomous indicator for student inflow for the same province where the university branch is located | Our computation on the dataset |
| *Student-teacher ratio* | Average number of enrolled students per academic professors in the university-field-province of destination | MIUR (2002-2011) |
| *Average number of ISI publications* | Average number of publications collected on WoS | ISI WoS (downloaded 2012) |
| *Average number of ISI citations* | Average number of citations of all the publications collected on ISI WoS | ISI WoS (downloaded 2012) |
| *Average impact factor* | Average impact factor of all the publications collected from WoS | ISI WoS + Journal Citation Reports (downloaded 2012) |
| *Average number of GS publications* | Average number of publications collected on Google Scholar | Google Scholar (downloaded 2015) |
| *Average number of GS citations* | Average number of citations of all the publications collected from Google Scholar | Google Scholar (downloaded 2015) |
| *Average number of Scopus publications* | Average number of publications collected from Scopus | Scopus (downloaded 2016) |
| *Average number of Scopus citations* | Average number of publications collected from Scopus | Scopus (downloaded 2016) |

*Note:* All variables refer to the 2003-2011 period.

## Table 2: Descriptive statistics for the main estimation sample

| Variables | Observations | Mean | Standard deviation |
|---|---|---|---|
| Student flows | 768,521 | 3.044849 | 31.47796 |
| Distance (km) | 768,521 | 5.75757 | 1.240214 |
| Sea | 768,521 | 0.248664 | 0.432239 |
| Sea x Distance | 768,521 | 1.592395 | 2.77474 |
| Province contiguity | 768,521 | 0.044164 | 0.20546 |
| Flow within same region | 768,521 | 0.06351 | 0.243877 |
| Within province | 768,521 | 0.008932 | 0.094088 |
| Student-teacher ratio (log) | 768,521 | 3.983975 | 1.6465 |
| Average number of ISI publications | 768,521 | -0.39793 | 1.994999 |
| Average number of ISI citations | 768,521 | 0.168984 | 2.789576 |
| Average impact factor | 768,521 | -0.32367 | 2.195421 |
| Average number of GS publications | 768,521 | 0.797862 | 0.860721 |
| Average number of GS citations | 768,521 | 2.718828 | 1.416716 |
| Average number of Scopus publications | 768,521 | 0.232759 | 1.951678 |
| Average number of Scopus citations | 768,521 | 1.56803 | 2.836342 |

## Table 3: Model with time-variant university-branches fixed effects

| Variables | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| *Distance (km)* | -0.983*** | -0.984*** | -0.984*** | -0.984*** | -0.984*** | -0.984*** | -0.984*** |
| | (0.059) | (0.059) | (0.059) | (0.059) | (0.059) | (0.059) | (0.059) |
| *Sea* | 1.351** | 1.353** | 1.353** | 1.353** | 1.353** | 1.353** | 1.353** |
| | (0.578) | (0.578) | (0.578) | (0.578) | (0.578) | (0.578) | (0.578) |
| *Sea × Distance* | -0.353*** | -0.354*** | -0.354*** | -0.354*** | -0.354*** | -0.354*** | -0.354*** |
| | (0.101) | (0.101) | (0.101) | (0.101) | (0.101) | (0.101) | (0.101) |
| *Province contiguity* | 1.113*** | 1.116*** | 1.116*** | 1.116*** | 1.116*** | 1.116*** | 1.116*** |
| | (0.111) | (0.111) | (0.111) | (0.111) | (0.111) | (0.111) | (0.111) |
| *Flow within same region* | 1.160*** | 1.160*** | 1.160*** | 1.160*** | 1.160*** | 1.160*** | 1.160*** |
| | (0.133) | (0.133) | (0.133) | (0.133) | (0.133) | (0.133) | (0.133) |
| *Within province* | -5.295*** | -5.302*** | -5.302*** | -5.302*** | -5.302*** | -5.302*** | -5.302*** |
| | (0.554) | (0.554) | (0.554) | (0.554) | (0.554) | (0.554) | (0.554) |
| *Student-teacher ratio* | 0.192*** | 0.193*** | 0.192*** | 0.194*** | 0.192*** | 0.192*** | 0.192*** |
| | (0.017) | (0.017) | (0.017) | (0.016) | (0.016) | (0.017) | (0.017) |
| *Average number of ISI publications* | 0.026*** | | | | | | |
| | (0.007) | | | | | | |
| *Average number of ISI citations* | | 0.013** | | | | | |
| | | (0.005) | | | | | |
| *Average impact factor* | | | 0.017*** | | | | |
| | | | (0.005) | | | | |
| *Average number of GS publications* | | | | 0.038** | | | |
| | | | | (0.015) | | | |
| *Average number of GS citations* | | | | | 0.034*** | | |
| | | | | | (0.009) | | |
| *Average number of Scopus publications* | | | | | | 0.059*** | |
| | | | | | | (0.01) | |
| *Average number of Scopus citations* | | | | | | | 0.026*** |
| | | | | | | | (0.006) |
| *Constant* | -1.508*** | 2.487*** | 2.504*** | 2.584*** | 2.593*** | 2.712*** | 2.564*** |
| | (0.498) | (0.614) | (0.612) | (0.625) | (0.622) | (0.606) | (0.609) |
| *Number of observations* | 757,617 | 768,521 | 768,521 | 768,521 | 768,521 | 768,521 | 768,521 |
| $R^2$ | 0.833 | 0.833 | 0.834 | 0.833 | 0.834 | 0.835 | 0.834 |

*Note:* The dependent variable is the number of students enrolled in a university branch, in a given subject-group coming from a given Italian province. All models also control for subject-group-year, branch-year and origin province-year fixed effects. Distance, the student-teacher ratio and research quality indicators are included in logarithms. Standard Errors in parenthesis. *, **, *** statistically significant at the 10%, 5% and 1% statistical levels, respectively.

## Table 4: Model with destination province-year fixed effects

| Variables | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| *Distance (km)* | -0.984*** | -0.984*** | -0.984*** | -0.984*** | -0.984*** | -0.984*** | -0.984*** |
| | (0.059) | (0.059) | (0.059) | (0.059) | (0.059) | (0.059) | (0.059) |
| *Sea* | 1.353** | 1.353** | 1.353** | 1.353** | 1.353** | 1.353** | 1.353** |
| | (0.578) | (0.578) | (0.578) | (0.578) | (0.578) | (0.578) | (0.578) |
| *Sea × Distance* | -0.354*** | -0.354*** | -0.354*** | -0.354*** | -0.354*** | -0.354*** | -0.354*** |
| | (0.101) | (0.101) | (0.101) | (0.101) | (0.101) | (0.101) | (0.101) |
| *Province contiguity* | 1.116*** | 1.116*** | 1.116*** | 1.116*** | 1.116*** | 1.116*** | 1.116*** |
| | (0.111) | (0.111) | (0.111) | (0.111) | (0.111) | (0.111) | (0.111) |
| *Flow within same region* | 1.160*** | 1.160*** | 1.160*** | 1.160*** | 1.160*** | 1.160*** | 1.160*** |
| | (0.133) | (0.133) | (0.133) | (0.133) | (0.133) | (0.133) | (0.133) |
| *Within province* | -5.302*** | -5.302*** | -5.302*** | -5.302*** | -5.302*** | -5.302*** | -5.302*** |
| | (0.554) | (0.554) | (0.554) | (0.554) | (0.554) | (0.554) | (0.554) |
| *Student-teacher ratio (log)* | 0.133*** | 0.131*** | 0.132*** | 0.125*** | 0.127*** | 0.142*** | 0.137*** |
| | (0.016) | (0.016) | (0.016) | (0.019) | (0.017) | (0.013) | (0.013) |
| *Average number of ISI publications* | 0.131*** | | | | | | |
| | (0.025) | | | | | | |
| *Average number of ISI citations* | | 0.085*** | | | | | |
| | | (0.016) | | | | | |
| *Average impact factor* | | | 0.106*** | | | | |
| | | | (0.021) | | | | |
| *Average number of GS publications* | | | | 0.160*** | | | |
| | | | | (0.045) | | | |
| *Average number of GS citations* | | | | | 0.146*** | | |
| | | | | | (0.031) | | |
| *Average number of Scopus publications* | | | | | | 0.145*** | |
| | | | | | | (0.030) | |
| *Average number of Scopus citations* | | | | | | | 0.083*** |
| | | | | | | | (0.018) |
| *Constant* | 0.611 | -1.832*** | -1.907*** | -1.782*** | 0.18 | -2.646 | -2.727*** |
| | (0.426) | (0.507) | (0.508) | (0.414) | (0.518) | (0.600) | (0.599) |
| *Number of observations* | 768,521 | 768,521 | 768,521 | 768,521 | 768,521 | 768,521 | 768,521 |
| $R^2$ | 0.732 | 0.734 | 0.739 | 0.706 | 0.724 | 0.746 | 0.739 |

*Note:* The dependent variable is the number of students enrolled in a university branch, in a given subject-group coming from a given Italian province. All models also control for subject-group-year, destination province-year and origin province-year fixed effects. Distance, the student-teacher ratio and research quality indicators are included in logarithms. Standard Errors in parenthesis. *, **, *** statistically significant at the 10%, 5% and 1% statistical levels, respectively.

**Table A1:** Linking teaching- and research-related subject-groups

| Subject-group (MIUR) | Subject-group (MIUR) | Scientific area (CUN) | Scientific area (CUN) |
|---|---|---|---|
| Natural sciences | 1 | 1+2* | Mathematics and informatics + Physics |
| Chemistry | 2 | 3 | Chemistry |
| Earth sciences and biology | 3 | 4 + 5* | Earth sciences + Biology |
| Medicine | 4 | 6 | Medicine |
| Agriculture | 7 | 7 | Agricultural and veterinary sciences |
| Architecture | 6 | 8 | Civil engineering and architecture |
| Engineering | 5 | 8 + 9* | Civil engineering and architecture + Industrial and information engineering |
| Literature | 11 | 10 + 11* | Antiquities, philology, literary studies, art history + History, philosophy, pedagogy and psychology |
| Languages | 12 | 10 + 11* | Antiquities, philology, literary studies, art history + History, philosophy, pedagogy and psychology |
| Teaching | 13 | 11 | History, philosophy, pedagogy and psychology |
| Psychology | 14 | 11 | History, philosophy, pedagogy and psychology |
| Law | 10 | 12 | Law |
| Economics and statistics | 8 | 13 | Economics and statistics |
| Political sciences | 9 | 14 | Political and social sciences |

*Note:* This table maps the teaching and research related classifications of subject groups used by the Italian Ministry of Education, University and Research (MIUR) and the National Research Council (CUN).

**Table B1:** Pairwise correlations between research quality indicators

| | Average number of ISI publications | Average number of ISI citations | Average impact factor | Average number of GS publications | Average number of GS citations | Average number of Scopus publications | Average number of Scopus citations |
|---|---|---|---|---|---|---|---|
| Average number of ISI publications | 1 | | | | | | |
| Average number of ISI citations | 0.8168 | 1 | | | | | |
| | [0.000] | | | | | | |
| Average impact factor | 0.9212 | 0.8835 | 1 | | | | |
| | [0.000] | [0.000] | | | | | |
| Average number of GS publications | 0.5157 | 0.4202 | 0.4821 | 1 | | | |
| | [0.000] | [0.000] | [0.000] | | | | |
| Average number of GS citations | 0.5207 | 0.518 | 0.5153 | 0.8542 | 1 | | |
| | [0.000] | [0.000] | [0.000] | [0.000] | | | |
| Average number of Scopus publications | 0.5909 | 0.5083 | 0.5737 | 0.5086 | 0.5162 | 1 | |
| | [0.000] | [0.000] | [0.000] | [0.000] | [0.000] | | |
| Average number of Scopus citations | 0.5491 | 0.5113 | 0.5491 | 0.4412 | 0.4957 | 0.8886 | 1 |
| | [0.000] | [0.000] | [0.000] | [0.000] | [0.000] | [0.000] | |

*Note:* The *P*-value is reported in brackets under each correlation coefficient.

**Table D1:** Model with time-variant university-branches fixed effects (before 2007)

| Variables | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| *Distance (km)* | -0.964*** | -0.965*** | -0.965*** | -0.965*** | -0.965*** | -0.965*** | -0.965*** |
| | (0.061) | (0.061) | (0.061) | (0.061) | (0.061) | (0.061) | (0.061) |
| *Sea* | 1.504** | 1.505** | 1.505** | 1.505** | 1.505** | 1.505** | 1.505** |
| | (0.603) | (0.603) | (0.603) | (0.603) | (0.603) | (0.603) | (0.603) |
| *Sea × Distance* | -0.380*** | -0.380*** | -0.380*** | -0.380*** | -0.380*** | -0.380*** | -0.380*** |
| | (0.106) | (0.106) | (0.106) | (0.106) | (0.106) | (0.106) | (0.106) |
| *Province contiguity* | 1.112*** | 1.114*** | 1.114*** | 1.114*** | 1.114*** | 1.114*** | 1.114*** |
| | (0.114) | (0.114) | (0.114) | (0.114) | (0.114) | (0.114) | (0.114) |
| *Flow within same region* | 1.185*** | 1.186*** | 1.186*** | 1.186*** | 1.186*** | 1.186*** | 1.186*** |
| | (0.135) | (0.135) | (0.135) | (0.135) | (0.135) | (0.135) | (0.135) |
| *Within province* | -5.121*** | -5.127*** | -5.127*** | -5.127*** | -5.127*** | -5.127*** | -5.127*** |
| | (0.571) | (0.571) | (0.571) | (0.571) | (0.571) | (0.571) | (0.571) |
| *Student-teacher ratio* | 0.209*** | 0.210*** | 0.210*** | 0.211*** | 0.209*** | 0.210*** | 0.210*** |
| | (0.019) | (0.019) | (0.019) | (0.018) | (0.018) | (0.019) | (0.019) |
| *Average number of ISI publications* | 0.026*** | | | | | | |
| | (0.008) | | | | | | |
| *Average number of ISI citations* | | 0.015*** | | | | | |
| | | (0.005) | | | | | |
| *Average impact factor* | | | 0.019*** | | | | |
| | | | (0.005) | | | | |
| *Average number of GS publications* | | | | 0.030** | | | |
| | | | | (0.014) | | | |
| *Average number of GS citations* | | | | | 0.027*** | | |
| | | | | | (0.008) | | |
| *Average number of Scopus publications* | | | | | | 0.046*** | |
| | | | | | | (0.01) | |
| *Average number of Scopus citations* | | | | | | | 0.019*** |
| | | | | | | | (0.006) |
| *Constant* | -2.862*** | 1.953*** | 1.963*** | 1.918*** | 1.899*** | 2.072*** | 1.955*** |
| | (0.388) | (0.380) | (0.378) | (0.380) | (0.380) | (0.367) | (0.377) |
| *Number of observations* | 375,924 | 382,476 | 382,476 | 382,476 | 382,476 | 382,476 | 382,476 |
| $R^2$ | 0.835 | 0.835 | 0.836 | 0.835 | 0.836 | 0.837 | 0.836 |

*Note:* The dependent variable is the number of students enrolled in a university branch, in a given subject-group coming from a given Italian province. All models also control for subject-group-year, branch-year and origin province-year fixed effects. Distance, the student-teacher ratio and research quality indicators are included in logarithms. Standard Errors in parenthesis. *, **, *** statistically significant at the 10%, 5% and 1% statistical levels, respectively.

**Table D2:** Model with time-variant university-branches fixed effects (after 2007)

| Variables | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| *Distance (km)* | -1.003*** | -1.006*** | -1.006*** | -1.006*** | -1.006*** | -1.006*** | -1.006*** |
| | (0.058) | (0.058) | (0.058) | (0.058) | (0.058) | (0.058) | 0.058 |
| *Sea* | 1.171** | 1.165** | 1.165** | 1.165** | 1.165** | 1.165** | 1.165** |
| | (0.575) | (0.574) | (0.574) | (0.574) | (0.574) | (0.574) | (0.574) |
| *Sea × Distance* | -0.321*** | -0.321*** | -0.321*** | -0.321*** | -0.321*** | -0.321*** | -0.321*** |
| | (0.100) | (0.100) | (0.100) | (0.100) | (0.100) | (0.100) | (0.100) |
| *Province contiguity* | 1.116*** | 1.118*** | 1.118*** | 1.118*** | 1.118*** | 1.118*** | 1.118*** |
| | (0.111) | (0.111) | (0.111) | (0.111) | (0.111) | (0.111) | (0.111) |
| *Flow within same region* | 1.128*** | 1.128*** | 1.128*** | 1.128*** | 1.128*** | 1.128*** | 1.128*** |
| | (0.135) | (0.135) | (0.135) | (0.135) | (0.135) | (0.135) | (0.135) |
| *Within province* | -5.484*** | -5.514*** | -5.514*** | -5.514*** | -5.514*** | -5.514*** | -5.514*** |
| | (0.547) | (0.546) | (0.546) | (0.546) | (0.546) | (0.546) | (0.546) |
| *Student-teacher ratio* | 0.169*** | 0.171*** | 0.170*** | 0.171*** | 0.170*** | 0.170*** | 0.168*** |
| | (0.015) | (0.016) | (0.015) | (0.015) | (0.015) | (0.014) | (0.014) |
| *Average number of ISI publications* | 0.025*** | | | | | | |
| | (0.009) | | | | | | |
| *Average number of ISI citations* | | 0.006 | | | | | |
| | | (0.009) | | | | | |
| *Average impact factor* | | | 0.013 | | | | |
| | | | (0.008) | | | | |
| *Average number of GS publications* | | | | 0.057* | | | |
| | | | | (0.032) | | | |
| *Average number of GS citations* | | | | | 0.050*** | | |
| | | | | | (0.016) | | |
| *Average number of Scopus publications* | | | | | | 0.086*** | |
| | | | | | | (0.014) | |
| *Average number of Scopus citations* | | | | | | | 0.042*** |
| | | | | | | | (0.010) |
| *Constant* | 2.366*** | 2.728*** | 2.769*** | 2.965*** | 2.956*** | 3.128*** | 2.949*** |
| | (0.425) | (0.603) | (0.600) | (0.637) | (0.621) | (0.603) | (0.591) |
| *Number of observations* | 379,725 | 386,045 | 386,045 | 386,045 | 386,045 | 386,045 | 386,045 |
| $R^2$ | 0.831 | 0.83 | 0.831 | 0.83 | 0.831 | 0.832 | 0.832 |

*Note:* The dependent variable is the number of students enrolled in a university branch, in a given subject-group coming from a given Italian province. All models also control for subject-group-year, branch-year and origin province-year fixed effects. Distance, the student-teacher ratio and research quality indicators are included in logarithms. Standard Errors in parenthesis. *, **, *** statistically significant at the 10%, 5% and 1% statistical level, respectively.

**Table D3:** Model with destination province-year fixed effects (before 2007)

| Variables | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| *Distance (km)* | -0.965*** | -0.965*** | -0.965*** | -0.965*** | -0.965*** | -0.965*** | -0.965*** |
| | (0.061) | (0.061) | (0.061) | (0.061) | (0.061) | (0.061) | (0.061) |
| *Sea* | 1.505** | 1.505** | 1.505** | 1.505** | 1.505** | 1.505** | 1.505** |
| | (0.603) | (0.603) | (0.603) | (0.603) | (0.603) | (0.603) | (0.603) |
| *Sea × Distance* | -0.380*** | -0.380*** | -0.380*** | -0.380*** | -0.380*** | -0.380*** | -0.380*** |
| | (0.106) | (0.106) | (0.106) | (0.106) | (0.106) | (0.106) | (0.106) |
| *Province contiguity* | 1.114*** | 1.114*** | 1.114*** | 1.114*** | 1.114*** | 1.114*** | 1.114*** |
| | (0.114) | (0.114) | (0.114) | (0.114) | (0.114) | (0.114) | (0.114) |
| *Flow within same region* | 1.186*** | 1.186*** | 1.186*** | 1.186*** | 1.186*** | 1.186*** | 1.186*** |
| | (0.135) | (0.135) | (0.135) | (0.135) | (0.135) | (0.135) | (0.135) |
| *Within province* | -5.127*** | -5.127*** | -5.127*** | -5.127*** | -5.127*** | -5.127*** | -5.127*** |
| | (0.571) | (0.571) | (0.571) | (0.571) | (0.571) | (0.571) | (0.571) |
| *Student-teacher ratio* | 0.134*** | 0.135*** | 0.136*** | 0.124*** | 0.125*** | 0.140*** | 0.139*** |
| | (0.016) | (0.016) | (0.016) | (0.019) | (0.018) | (0.014) | (0.013) |
| *Average number of ISI publications* | 0.127*** | | | | | | |
| | (0.020) | | | | | | |
| *Average number of ISI citations* | | 0.083*** | | | | | |
| | | (0.013) | | | | | |
| *Average impact factor* | | | 0.105*** | | | | |
| | | | (0.017) | | | | |
| *Average number of GS publications* | | | | 0.125*** | | | |
| | | | | (0.038) | | | |
| *Average number of GS citations* | | | | | 0.123*** | | |
| | | | | | (0.025) | | |
| *Average number of Scopus publications* | | | | | | 0.134*** | |
| | | | | | | (0.025) | |
| *Average number of Scopus citations* | | | | | | | 0.080*** |
| | | | | | | | (0.015) |
| *Constant* | 2.451*** | 2.219*** | 0.197 | 2.345*** | 2.024*** | 2.276*** | 2.114*** |
| | (0.426) | (0.459) | (0.425) | (0.462) | (0.447) | (0.455) | (0.452) |
| *Number of observations* | 382,476 | 382,476 | 382,476 | 382,476 | 382,476 | 382,476 | 382,476 |
| $R^2$ | 0.731 | 0.739 | 0.743 | 0.701 | 0.72 | 0.744 | 0.741 |

*Note:* The dependent variable is the number of students enrolled in a university branch, in a given subject-group coming from a given Italian province. All models also control for subject-group-year, destination province-year and origin province-year fixed effects. Distance, the student-teacher ratio and research quality indicators are included in logarithms. Standard Errors in parenthesis. *, **, *** statistically significant at the 10%, 5% and 1% statistical level, respectively.

**Table D4:** Model with destination province-year fixed effects (after 2007)

| Variables | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| *Distance (km)* | -1.006*** | -1.006*** | -1.006*** | -1.006*** | -1.006*** | -1.006*** | -1.006*** |
| | (0.058) | (0.058) | (0.058) | (0.058) | (0.058) | (0.058) | (0.058) |
| *Sea* | 1.165** | 1.165** | 1.165** | 1.165** | 1.165** | 1.165** | 1.165** |
| | (0.574) | (0.574) | (0.574) | (0.574) | (0.574) | (0.574) | (0.574) |
| *Sea × Distance* | -0.321*** | -0.321*** | -0.321*** | -0.321*** | -0.321*** | -0.321*** | -0.321*** |
| | (0.100) | (0.100) | (0.100) | (0.100) | (0.100) | (0.100) | (0.100) |
| *Province contiguity* | 1.118*** | 1.118*** | 1.118*** | 1.118*** | 1.118*** | 1.118*** | 1.118*** |
| | (0.111) | (0.111) | (0.111) | (0.111) | (0.111) | (0.111) | (0.111) |
| *Flow within same region* | 1.128*** | 1.128*** | 1.128*** | 1.128*** | 1.128*** | 1.128*** | 1.128*** |
| | (0.135) | (0.135) | (0.135) | (0.135) | (0.135) | (0.135) | (0.135) |
| *Within province* | -5.514*** | -5.514*** | -5.514*** | -5.514*** | -5.514*** | -5.514*** | -5.514*** |
| | (0.546) | (0.546) | (0.546) | (0.546) | (0.546) | (0.546) | (0.546) |
| *Student-teacher ratio* | 0.130*** | 0.125*** | 0.126*** | 0.127*** | 0.131*** | 0.145*** | 0.136*** |
| | (0.017) | (0.018) | (0.018) | (0.018) | (0.016) | (0.013) | (0.015) |
| *Average number of ISI publications* | 0.138*** | | | | | | |
| | (0.037) | | | | | | |
| *Average number of ISI citations* | | 0.090*** | | | | | |
| | | (0.028) | | | | | |
| *Average impact factor* | | | 0.108*** | | | | |
| | | | (0.032) | | | | |
| *Average number of GS publications* | | | | 0.211*** | | | |
| | | | | (0.055) | | | |
| *Average number of GS citations* | | | | | 0.189*** | | |
| | | | | | (0.042) | | |
| *Average number of Scopus publications* | | | | | | 0.165*** | |
| | | | | | | (0.042) | |
| *Average number of Scopus citations* | | | | | | | 0.087*** |
| | | | | | | | (0.025) |
| *Constant* | 3.593*** | 3.643*** | 3.529*** | 3.109*** | 3.038*** | 3.204*** | 3.316*** |
| | (0.457) | (0.469) | (0.456) | (0.440) | (0.439) | (0.439) | (0.442) |
| *Number of observations* | 386,045 | 386,045 | 386,045 | 386,045 | 386,045 | 386,045 | 386,045 |
| $R^2$ | 0.734 | 0.726 | 0.731 | 0.714 | 0.729 | 0.747 | 0.736 |

*Note:* The dependent variable is the number of students enrolled in a university branch, in a given subject-group coming from a given Italian province. All models also control for subject-group-year, branch-year and destination province-year fixed effects. Distance, the student-teacher ratio and research quality indicators are included in logarithms. Standard Errors in parenthesis. *, **, *** statistically significant at the 10%, 5% and 1% statistical level, respectively.

**Table D5:** Models of research quality attractiveness (by geographic area of destination)

| Research quality indicators | Statistics | Equation (9) (time-variant university-branches FEs) | | | | Equation (10) (destination province-year FEs) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | North | Centre | South | Islands | North | Centre | South | Islands |
| | | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Average number of ISI publications | β | 0.038*** | -0.005 | 0.057*** | -0.018 | 0.086*** | 0.226*** | 0.124*** | 0.027 |
| | SE | 0.01 | 0.013 | -0.015 | 0.011 | 0.017 | 0.028 | 0.027 | 0.02 |
| | $R^2$ | 0.853 | 0.892 | 0.809 | 0.917 | 0.78 | 0.719 | 0.749 | 0.865 |
| Average number of ISI citations | β | 0.035*** | 0.003 | 0.014 | -0.023*** | 0.064*** | 0.166*** | 0.070*** | 0.013 |
| | SE | 0.007 | 0.007 | 0.013 | 0.008 | 0.013 | 0.018 | 0.017 | 0.013 |
| | $R^2$ | 0.855 | 0.891 | 0.806 | 0.917 | 0.786 | 0.744 | 0.741 | 0.864 |
| Average impact factor | β | 0.038*** | -0.008** | 0.028** | -0.024*** | 0.076*** | 0.204*** | 0.086*** | 0.02 |
| | SE | 0.007 | 0.009 | -0.012 | 0.008 | 0.017 | 0.021 | 0.022 | 0.014 |
| | $R^2$ | 0.855 | 0.892 | 0.807 | 0.917 | 0.786 | 0.764 | 0.744 | 0.864 |
| Average number of GS publications | β | 0.015 | 0.026 | 0.101*** | 0.118*** | 0.006 | 0.229*** | 0.198*** | 0.116*** |
| | SE | 0.021 | 0.024 | 0.037 | 0.029 | 0.02 | 0.057 | 0.064 | 0.028 |
| | $R^2$ | 0.852 | 0.891 | 0.81 | 0.916 | 0.767 | 0.656 | 0.731 | 0.862 |
| Average number of GS citations | β | 0.014 | 0.02 | 0.074*** | 0.050*** | 0.053*** | 0.194*** | 0.148*** | 0.129*** |
| | SE | 0.015 | 0.014 | 0.018 | 0.013 | 0.018 | 0.048 | 0.035 | 0.048 |
| | $R^2$ | 0.852 | 0.891 | 0.811 | 0.918 | 0.77 | 0.689 | 0.74 | 0.883 |
| Average number of Scopus publications | β | 0.056*** | -0.006 | 0.088*** | 0.005 | 0.057* | 0.224*** | 0.147*** | 0.119** |
| | SE | 0.01 | 0.025 | 0.014 | 0.034 | 0.031 | 0.046 | 0.033 | 0.053 |
| | $R^2$ | 0.853 | 0.892 | 0.814 | 0.917 | 0.778 | 0.732 | 0.769 | 0.874 |
| Average number of Scopus citations | β | 0.033*** | -0.007 | 0.034*** | 0.014 | 0.032*** | 0.162*** | 0.075*** | 0.108** |
| | SE | 0.005 | 0.025 | 0.007 | 0.018 | 0.01 | 0.029 | 0.023 | 0.049 |
| | $R^2$ | 0.853 | 0.892 | .0.810 | 0.917 | 0.777 | 0.74 | 0.753 | 0.886 |
| Observations | N | 329,421 | 177,816 | 158,919 | 62,990 | 329,421 | 177,816 | 158,919 | 62,990 |

*Note:* The dependent variable is the number of students enrolled in a university branch, in a given subject-group coming from a given Italian province. The table reports only the coefficients of the research quality indicators. All models also control for the covariates included in equations (9) and (10). *, **, *** statistically significant at the 10%, 5% and 1% statistical level, respectively.