Università degli Studi di Milano

Graduate School in Social, Economic and Political Sciences -Department of Economics, Management and Quantitative Methods

PhD in Economics [SECS-P/01] - XXV Cycle

Efficiency and Scientific Productivity in Italian Universities

Stefano Verzillo

Matr. n° R08479

PhD Supervisors:

Prof. Daniele Checchi (Università degli Studi di Milano)

Prof. Gianni De Fraja (University of Nottingham)

PhD Coordinator:

Prof. Michele Santoni (Università degli Studi di Milano)

Academic Year: 2011-2012

Acknowledgments

Academics' research effort is usually driven by new and (hopefully) exciting questions. Collecting evidence to answer these questions and translating them into papers is the subsequent 'hard' work of researchers.

What I'm actually sure after three years of Ph.D. is that when I start doing my research I never reach a definitive end. Questions often raise new other questions, robustness checks could always be completed and finally ideas are followed by other ideas. The hardest part of my doctoral studying has been to fix a point, at a certain moment, and then conclude the research papers, writing all evidences and results, while also being aware of the fact that the research has only been started.

So, having provisionally finished my Ph.D. thesis, I would like to especially thank all those that helped me during these years of studying and research. I am most grateful to my doctoral colleagues (Laura Casi, Endalamaw Bati Bayou and William Seitz) for many fruitful discussions and to the Ph.D director, Professor Michele Santoni, for his support.

As part of the postgraduate program I also spent an academic year at the Economics Department of Leicester University (a.y. 2010/2011). I would like to especially thank my colleagues and friends in the UK Lorenzo, Paolo, Diego, Monica, Tanja, Anita, Tenday, Andry, Francesca and James for their support and the nice time spent together in the United Kingdom.

Then I am grateful to several people at the Statistics Department of University of Milan-Bicocca and at the Department of Economics, Business and Statistics at University of Milan (Paolo Berta, Benedetta Trivellato, Mario Mezzanzanica, Piergiorgio Lovaglio, Stefano Iacus, Silvia Salini, Carlo Fiorio, Massimiliano Bratti and Antonio Filippin) for their suggestions and useful comments in the early, middle and final stages of this work.

I also wrote the third chapter of this thesis together with my supervisor Prof. Daniele Checchi. All these years, he has been an excellent supervisor, teaching me a lot about how academic research really should be done in this field. I'm also grateful for the co-supervision of Prof. Gianni De Fraja (at University of Leicester), a really sharp and refined economist who is a great pleasure to discuss with.

Then a special mention has to be done for Ing. Matteo Fontana who helped me in solving several problems with data-matching techniques at Crisp - Interuniversity Research Centre on Public Services at University of Milan-Bicocca.

The Italian Ministry of Education (Miur), Crisp and the Cariplo Foundation provided financial support at different stages of my last three years of work. This support is gratefully acknowledged.

Finally I would like to really thank my parents for their support, and especially my wife Beatrice and my “little” son, Giacomo, for making the life worth living during these last years I’ve spent studying and working on this dissertation. A special thought goes also to Daniela Usai.

Milano, December 2012

Stefano Verzillo

Contents

1 Introduction

Efficiency in Higher Education Systems is an issue of crucial importance these days considering the strong dependence of this sector on public funds (especially in the case of Italy where the share of public funds in 2010 was 91% of the overall expenditure). Indeed in the context of the current economic crisis the evaluation of efficiency of education provision is a primary target for governments from a social viewpoint. Several studies in the scientific literature started in the early 1990s to explore the issue of efficiency in the Higher Education Sector by focusing on different particular aspects.

In this context, this thesis seeks to fill a significant gap in the literature on the Italian higher education system, by looking at issues of efficiency and quality as described below. The thesis consists of three essays, one of which (chapter 3) is a joint paper with my supervisor Daniele Checchi (University of Milan).

Hence the first paper is strictly related to the estimation of the overall efficiency of Italian Universities during the last decade using appropriate recently introduced (Tsionas, 2002; Greene, 2005) econometric techniques. Efficiency is measured by taking into account the multidimensional nature of academic institutions which -by their nature- involved teaching, research and administrative activities as their primary dimensions. Going deeply into the issue of academics' efficiency, we should note that the main human capital stock of inputs for both the teaching and research dimensions consists of the academic body.

Quality, competences and professionalism of professors are leading "determinants" of teaching and research of our universities. Therefore the selection mechanisms of academics and their recruitment rules gather a great importance in this context. To this purpose we notice that a decentralization of academic selection procedures was introduced in Italy in 1998 with the "Berlinguer reform" act.

The second and the third paper of this thesis focus especially on the evaluation of the quality of those individuals who succeeded in competitive examinations to become associate, and full professors, and on the incentives to produce international research associated with different recruitment mechanisms (national versus local).

In order to build the empirical basis for our research, we have expressly assembled an ad hoc database by creating an algorithm to match journal-publications of the web version of ISI-Web of Knowledge with the institutional database of

Italian scientists (provided by the Italian Ministry of University and Research - Miur) and with the Journal Citation Index (JCI 2010). This way we set up a database of academic careers of the entire population of Italian university professors over the last two decades in terms of their international research products (with measures on impact, quantity and notoriety). ISI records with associated bibliometric indicators provide us with comparable international measures of the research performance of the academics involved.

Starting from the 1990s, we found in the scientific literature several studies on the research productivity of academics. Examples are Levin and Stephan (1991) who studied the relationship between publishing productivity and age of Ph.D. scientists using individual US data, and Noser, Manakyan and Tanner (1996) who focused the attention on the relationship between research productivity and the perceived teaching effectiveness of economics scientists. Furthermore, a growing literature arose with the advent of available bibliometric sources of information that addressed the possibility of evaluating institutions, departments and individuals with objective quantitative measures. Indeed empirical evidences showed an increase of the publishing trend in the last two decades (the so-called “secular progress of knowledge” of J.Mincer, 1974) and a large degree of heterogeneity in bibliometric measures across different academic disciplines (Angelucci, M. et al., 2010).

Over the last decade a large literature has documented pros. and cons. of applying bibliometric indicators to evaluation exercises (Falagas et al., 2008; Bakkalbasi et al., 2006). The role of the impact factor (Seglen, P.O., 1997), ownerships of the bibliometric sources (private companies and open source services), differences in disciplinary coverages (Klavan and Boyak, 2007), correlations between bibliometric measures of different sources (Archambault et al., 2009) and the use of citations statistics (Adler et al., 2009) are some of the most debated topics in this field. In what follows, considering that bibliometric indicators are -at this moment- the only available, objective, standard accepted measures of international research outcomes of Italian academics over the last twenty years, we focussed on the effects on these international research outcomes of decentralizing the recruitment of professors from the national to the university level.

To date -at my knowledge- no studies in the scientific literature deal with the effects of the 1998 reform (which decentralized by law academic recruitment selections) on the scientists' research productivity in Italy, except for Labartino et al. (2009) where the main goals were both to test the presence of misbehaviors such as familism and nepotism in the Italian academia (evidenced by the differences in the distributions of surnames with respect to the provincial population) and the possible relations of these phenomena with different levels of civic capital.

All the three chapters are written as single working papers to be read individually by the readers. Therefore it is possible to note a certain degree of overlapping that was clearly unavoidable. In the remaining part of this introduction I briefly introduce the topic of each paper, the econometric techniques used, and I summarize the main findings of each essay.

The Italian University System: A Technical Efficiency Analysis

In the first paper we estimates the efficiency scores of Italian universities over a panel of four years from 2005 to 2008. The proposed approach considers a multiple-input and multiple-output production function (Baumol et al., 1982) where teaching, management and research are considered as different dimensions of the same productive technical process.

The production function we used has a trans-logarithmic functional form to take into account the multidimensionality of the higher education system (multiple are the 'missions' of an Italian University: knowledge transfer, research effort, social mission...). From an econometric viewpoint, a recent version of the stochastic frontiers model with random coefficients (Tsionas, 2002; Greene, 2005) allows to simultaneously consider the heterogeneity of academic institutions and the institutional inter-temporal dynamic. The proposed model incorporates at the same level the trans-logarithmic input-output distance function and some background control variables. An institution-specific variable to disentangle the effect of subject-mix composition of universities on their research outputs, a regional-specific covariate and other institutional variables are introduced in order to control for background and territorial differences between institutions.

Usually efficiency studies in higher education are estimated through indirect measures of research outputs (research amount of funds/grants) due to the lack of more appropriate data. A novelty with respect to the current literature is indeed the dataset we used: particularly interesting is the availability of direct measures of research outputs (taking into account both quality and quantity dimensions of the research effort) at the institutional level.

Our dataset is assembled with official data published by the Italian Ministry of Education and Research (MIUR) and direct measures of research outputs thanks to the web version of ISI Web Of Knowledge and its Journal Citations Report. This study is one of the first panel studies on efficiency scores in the Italian case. Some methodological implications such as the bias of estimates due to grade inflation in quality measures of both teaching inputs and outputs are widely discussed in the paper. Grade inflation is persistent in Italy and both empirical evidences (Modica, 2008; Bagues, et al., 2008; Bratti et al. 2010) and its theoretical framework (Tampieri, A, 2012) are provided by the recent literature. We show that efficiency scores are strongly biased by the inclusion of teaching input measures weighted by grades (as inclusive of quality) instead of quantity only.

The empirical results of this study suggest that the levels of inputs do not vary so much across institutions with different efficiency but what varies is their outputs composition. The most efficient institutions tend to be the largest producers of both graduates and postgraduates but not absolutely the largest producers by quantity and quality of research products.

Selecting University Professors in Italy: much 'ado' about nothing[1]

The second and third papers are both focussed on the relationship between the recruitment mechanisms of academic professors and their research productivity (Allesina, S., 2011; Checchi, D., 1999). Particularly in the second chapter we considered whether the introduction of local selection procedures after the 1998 reform influences the research quality of selected academics (measured by ISI Web of Knowledge records). We tested the presence (or not) of a negative effect of local selections on the productivity of Italian professors at the time of their selection (both for associates and full professorships).

As a first result focussing our analysis on the quality level (measured by bibliometric indicators of quantity and impact of publications) of the promoted individuals relative to the corresponding losers, we found no clear picture of a general worsening, as previously noticed by the literature (Labartino at al.

[1] Joint with Daniele Checchi

2011).

The major finding of this study is that no average effect is associated with the decentralizing reform over all the disciplinary areas. Nevertheless, when the world is restricted to 'bibliometrics disciplines' only a negative -even if poor-significant effect arises. This is due to the large heterogeneity of data across disciplines: strong negative effects of local selections in the case of few sciences (Chemistry, Biology, Physics and Industrial Engineering) are combined with null effects over the remaining disciplines. This is verified with some differences in discplines both for associate and full professorships. However, what is clearly evident is the increase invariability in the quality of selected professors due to the reform impact.

Hence we asked whether this increase in the quality variability of promoted professors might be the result of increased polarisation in the criteria adopted by the selecting committees. Results are again contrasting: candidates with a higher scientific productivity are less likely to be selected by a department with a higher (on average) productivity, but this effect is attenuated after the reform. This implies that one of the reform's effects is an increase in the polarization of department/candidates behaviours. An opposite effect is associated with the reform when considering measures of research impact and notoriety of selected candidates.

Decentralized Academic Selection Mechanisms: Opportunity or Parochialism?

The third chapter deals with the econometric measure of individual research incentives associated with different recruitment mechanisms in Italy. We focused primarily on the differences between individual research patterns (both in terms of quantity and impact) before and after the 1998 decentralizing reform for scientists recruited as both associates and full professors in a "quasi-experimental" research framework. The underlying idea is to identify two groups of comparable scientists (with balanced levels of productivity, same discipline and same position at time of selection) who were selected one before and the other after the introduction of the 1998 reform.

Methodologically we applied two different strategies: a standard propensity score matching approach and a recently proposed non-parametric matching algorithm called Coarsened Exact Matching-CEM (Iacus, S., 2009).

The first methodology is used as benchmark and it is based on observational studies literature where propensity score can be used in different manners: included as an adjustment covariate or used in a quintile regression framework or finally it can also be the matching factor in indentifying treated and controls groups on which apply treatment models (Kleinman and Horton, 2009; Hedeker and Gibbons 1997). On the other hand, as our alternative empirical strategy we use CEM as matching algorithm. It allows to obtain balanced groups of treated (selected after the reform) and controls (before the reform) with respect to the available research productivity outcomes at the time of selection. The main advantage of this strategy relies on its non-parametric nature. The multidimensionality of the original data is preserved (while propensity score reduces it to an one-dimension score) and the balance of the two groups when tested performs really well. Balancing is set at the first stage and then the number of matching units is a consequence.

Results suggest that the story is different for each discipline, and that they cannot be generalized for the entire Italian academia. However, negative effects on impact research outcomes and slopes are identified with both strategies for people selected after decentralization, proving that lower incentives for publishing in international top-level journals are associated with the local selections in the case of both associate and full professors.

By differentiating the model by research areas, it is shown that negative incentives are associated with disciplines such as Math, Earth sciences, Medicine, Veterinary and Agricultural Sciences. The identified effects are robust on different checks.

Results are consistent with previous findings also in this case. Stated differently, there is evidence that, on average, scientists become less productive after the local reform in terms of the impact of their research and in some cases in terms of the quantity of their research production.

References

[1] Adler R., J. Ewing J., Taylor P. (2009), "Citation statistics (with discussion)" Statistical Science, 24, 1-28;

[2] Allesina, S. (2011), "Measuring nepotism through shared last names: the case of Italian academia", PLoS one 6(8), e21160;

[3] Angelucci, M., G. De Giorgi, M. Rangel and Rasul I. (2010), "Family networks and school enrollment: evidence from a randomized social experiment", Journal of Public Economics 94(3-4), 197–221;

[4] Archambault E., Campbell D., Gingras Y., Larivire V. (2009), "Comparing Bibliometric Statistics Obtained from the Web of Science and Scopus", Journal of the American Society for Informatin Science and Technology, 60(7);

[5] Bagues, M, Labini, M.,S., Zynovyeva, N., (2008), "Differential Grading Standards and University Funding: Evidence from Italy", working paper series Capital Humano y Empleo – Santander;

[6] Baumol W. J., Panzar, J. C. and Willig, R. D. (1982),"Contestable Markets and the Theory of Industry Structure", Harcourt Brace Jovanovich, San Diego;

[7] Bratti, M., Broccolini, C., Staffolani, S. (2010), "Higher education reform, student time allocation and academic performance in Italy: Evidence from a Faculty of Economics", Rivista Italiana degli Economisti, SIE - Societa' Italiana degli Economisti (I), vol. 15(2), pages 275-304;

[8] Checchi, D. (1999), "Tenure. an appraisal of a national selection process for associate professorship", Giornale degli Economisti 58(2), 137–181;

[9] Durante, R., Labartino G., Perotti R., and Tabellini G. (2009), "Academic dynasties", Bocconi University Working Paper;

[10] Fox, M. F. (1983), "Publication Productivity Among Scientists: A Critical Review", Social Studies of Science, 13, 283-305;

[11] Hedeker, D., Gibbons, R.D., (2007), "Application of Random-Effects Pattern-Mixture Models for Missing Data in Longitudinal Studies", Psychological Methods, Vol. 2, No. 1, 64-78;

[12] Iacus, S. M., King G., Porro, G. (2009), "CEM: Coarsened Exact Matching Software", Journal of Statistical Software 30(9), http://gking.harvard.edu/cem;

[13] Klavan and Boyak, (2007) "Is there a Convergent Structure of Science? A Comparison of Maps using the ISI and Scopus Databases1", 11th International Conference of the International Society for Scientometrics and Informetrics meeting in Madrid, Spain, June 2007;

[14] Kleinman, K., Horton, N.J., (2009), "SAS and R: Data Management, Statistical Analysis, and Graphics", Ed. Chapman and Hall/CRC;

[15] Labartino G., Perotti R., Durante R. (2011), "Academic Dynasties: Decentralization and Familism in the Italian Academia", NBER Working Paper No. 17.572;

[16] Levin S., Stephan P. (1991), "Research Productivity Over The Life Cycle: Evidence for Academic Scientists", The American Economic Review, Vol. 81, Issue 1, 114-132;

[17] Mincer, J. (1974), "Schooling, Experience and Earnings", New York: Columbia University Press (for the National Bureau of Economic Research);

[18] Modica, S. (2008), "I voti di laurea non sono normali" http://www.lavoce.info/articoli/pagina1000421-351.html.it;

[19] Noser T., Manakyan H. and Tanner J. (1996), "Research Productivity and Perceived Teaching Effectiveness: A Survey of Economics Faculty", Research in Higher Education, Vol. 37 No. 3;

[20] Seglen P.O. (1997), "Why the impact factor should not be used for evaluating research" British Journal of Medicine, 314, 498-502;

[21] Tampieri, A. (2012), "Grade Inflation, Students Social Background and String-Pulling", working paper Università degli Studi di Bologna;

[22] Tsionas, E. G., (2002), "Stochastic frontier models with random coefficients", Journal of Applied Econometrics, 17, pp. 127–47;

2 The Italian University System: A Technical Efficiency Analysis

Abstract

The aim of this paper is to estimate efficiency scores of Italian universities using a recent evolution of the stochastic frontiers approach. We introduce a recently random coefficients version of the model (Tsionas, 2002; Greene, 2005), which takes into account the heterogeneity of the academic institutions and the inter-temporal dynamic. The production function is a multiple-input and multiple-output function (Baumol et al., 1982) where teaching and research are both considered as different dimensions of the same productive technical process.
A trans-log stochastic frontier production function model is estimated over a four-years period (2005- 2008) using an integrated data-set composed by official data published by the Italian Ministry of Education and Research (MIUR) and direct measures of research outputs. This study is one of the first panel studies on efficiency scores in the Italian case and it allows to measure the efficiency using direct measures of research outputs (taking into account both quality and quantity dimensions): the ISI – Thomson dataset of Philadelphia. One of the most common drawbacks in literature is that the estimation of the efficiency scores is fitted through indirect measures of research outputs because of the lack of available informations (amount of money collected from the public/private sector, number of research projects accepted/financed by the National Ministry of Education and Research etc…). Unfortunately the examination of the effects of quantity-quality measures of teaching inputs and outputs on efficiencies suggests the presence of a persistent grade-inflation in the Italian system. Grade inflation mainly affects south and middle regions Universities.

JEL Classification: I23 C01 C33 D24 Keywords: stochastic frontiers, technical efficiency, ISI Web of Knowledge

2.1 Introduction

The higher education sector in countries is commonly sustained by public funding. For instance, the amount of private sector expenditure for educational institutions in 2007 was 8,9% in Italy, 10,8% in the EU area and 17,4% in the OECD countries (Education at a Glance 2010). This shows the dependence of the Higher Education Institutions (HEIs) on public sector, especially in the case of Italy.

In this context, accountability has a specific and crucial importance, and measurements of the efficiency of education provision are a primary concern for each government from a social point of view. The aim of this paper is to measure the efficiency of Higher Education Decision Making Units (DMUs) by estimating efficiency scores, which can be used as a support in the political decision-making process. The recent financial crisis has determined a situation of austerity in terms of public spending, so that publicly funded sectors are facing with tight budgets. The University Reform Law recently approved (14th January, 2011) as an attempt to address some of the structural problems of the Italian system and one of its declared aims is the introduction of a "meritocratic system" were all the academic institutions are evaluated and financed with respect to their management capability and efficiency levels (D.L. 240/2010, Tit. I, art.1, co. 4). It has also been suggested that there is scope for absorbing part of the public fundings cuts by increased in the efficiency of the institutions (Mandelson, 2009). The research effort of this work would be finalized to find out some robust results that could be used by policy-makers.

Efficiency evaluation has been a topic of particular interest in economic literature over the last fifty years, which has been applied to different economic fields; for example, health (Thanassoulis, Boussfofiane and Dyson, 1995) and banking sectors (Golany and Storbeck, 1999). Despite this, the higher education sector (HES) is understudied, probably because of its non-profit nature and makes it difficult to study the efficiency. The view of universities as producers of graduates has been overhauled with the pioneering paper of Baumol, Panzar and Willing (1982). After that the HES was characterized by the presence of a multi-input and multi-output structure recognizing the multi-dimensional nature; it produces multiple outputs (e.g. research, graduates, patents, papers) from multiple inputs (e.g. students, research effort, fees, funds and grants. . .) with a process that can be modelled by a specific production function.

The international literature reports various studies about technical efficiency of educational systems (at all levels: high-schools or universities) in some different countries (USA and UK above all), but -according to my knowledge- only few studies in the Italian case (in both the two sub-sectors). The cornerstones of the efficiency studies in HES are papers of Athanassopoulos and Shale (1997) and J. Johnes application of the data envelopment analysis (DEA) at higher education performances (Izadi, Johnes, Oskrochi, Crouchley, 2002; Johnes, 2006). Other pioneering researches are related to the measure of departmental efficiency (Gimenez, J.L. Martinez, 2006), the student's study efficiency -considering their optimal time-allocation- (Dolton, Marcenaro, Navarro, 2003) and the aggregate efficiency of HEIs from an individual student's perspective, where each student is supposed to be a DMUs himself (Thanassoulis, 2001; Waldo, 2009).

Only a couple of papers are related to the measurement of technical efficiency (only using a cost function approach) through Stochastic Frontier Analysis (SFA) related to Italian universities (Agasisti, Johnes, 2010); a few others used Data Envelopment Analysis (DEA) instead of SFA approach (Agasisti, Johnes, 2009) again with a cost related perspective. The aim of the present paper is to measure the technical efficiency of the Italian HES by applying a recent methodological extension of the SFA (with random parameters) to a multi-input/multi output distance production function. The panel data-set is relative to all the Italian Universities between academic years 2005/2006 and 2008/2009 and it is used to estimate a stochastic frontier trans-log model where parameters are allowed to vary across institutions and over time. This random version of SFA approach is introduced in the work of Tsionas (2002) and Greene (2005) and is useful for separating technical efficiency from technological differences across universities. Usually in stochastic frontier framework this restrictive assumption is supposed: all universities must share exactly the same technological production function; it's a sort of "homogeneity assumption" in their accounting behaviour.

Nevertheless it is well known that each academic institution's production function may have a different functional form due to the different features of the institution itself. Therefore, we have to consider universities as a complex system in which differentiation is the key concept to understanding the global phenomena. The higher education sector in Italy is considered too often as a monolithic and homogeneous framework where universities are considered equal, with same management system, same level of teaching quality and sometimes same quality of research outputs. On the contrary each university has different excellences in teaching and research areas, different type of students and are settled in a different regional labour-market context. For these reasons, in our paper, we relax this hypothesis and we try to consider as well as possible the heterogeneity that it's own of the HES' nature.

Only recently few empirical studies of the HEIs sector have used frontier methods, instead of DEA (the most commonly used tools), to estimate models that provide measures of institutional efficiency (Johnes, 1997 and 2010; Agasisti, Johnes, 2010; Stevens, 2005). These studies are mostly focused on the UK Higher Education System.

An unusual aspect of the work is represented by the direct measure of research output we used. In literature a lot of different measures of research outputs have been used, but they are often not properly "direct". The amount of grants or funds received by a department (Agasisti and Johnes, 2008) or the number of research project financed by the government (Athanassopoulos and Shale, 1997) are examples of these "indirect" measures of the quality of a university's research. Properly grants could represent an input of the research process, and should not be used as a measure of the research output (Johnes and Johnes, 1993). In order to solve this problem, our paper uses the ISI – Thomson data-set for measuring the number of high-level research publications and papers that are produced by each Italian University in each of the selected years, while we use research grants (numbers of research projects funded) for measuring the research input associated.

Then our basic research question is: which methodology can best be employed to measure efficiency in HEIs taking into account both the multidimesnionality and heterogeneity of the input-output HE structure? Does it allow for institutional variability? In order to answer this research questiosn an analysis combining a translog distance function with multiple inputs and outputs with a RPM SFA is proposed. This is new to the literature and is particularly relevant because of the quality of data available for the study. Direct measures of quality and quantity of research output are available for the Italian system and this work tends to be the first attempt in the estimation of a complex -but more realistic-model of technical efficiency. As a result of the theoretically suggested inclusion of quality weights into teaching inputs and outputs we register strong bias due to grade-inflation. Italian institutional marks can not be considered as realistic metrics of students' quality.

The paper is structured as follows. The next section presents some brief information about the structure and the growing attention on evaluation processes in the Italian university system. The third section provides a wide explanation of the integrated data set used for the analysis and of the different sources that composed it. The fourth section presents in details, the random parameters stochastic frontier model with its advantages and drawbacks. Results are presented in section five with a plausible explanation of their meaning. Section six discuss grade-inflation issue and it's effect on efficiency rankings we found. Conclusions regarding efficiency of the Italian Universities are drawn in the last section.

2.2 The Italian University System

The Italian HE system is characterized by a strong central government given by the State since the beginning of its modern life (Casati Act, 1859). During the XX century the central education authority administrates the system in both financial issues and teaching aspects.

The autonomy of the universities, introduced by the Constituent law (1948), and the Legal Value of the degree title (introduced in 1973) were substantially undervalued since the end of 80's. The Italian system does not provide a binary division between vocational and academic institutions, typical of most of the European and USA countries, but it is composed by universities with same aims and structure that only differ in their sizes and locations (more details on the administrative evolution of Italian Higher Education System in Marra, A., 2009).

The most common feature of HES in developed countries is an heterogeneous systems composed by a double quality levels of education (academic and vocational), as in the case of the United Kingdom, Germany (Universitaeten and Fachhochschule) or France (Universities and Grandes Ecoles). Starting from the beginning of the 90's the financial issue of academic institutions increases in freedom of managing the financial funding (Financial law, 1993), which it started to be totally under their own responsibility. Actually the central authority has fixed only a few restrictions (e.g. 90% maximum proportion of staff's income expenditure and 20% maximum amount of tuition fees with respect to Ordinary Financial Fund - FFO) that have to be observed. On the other hand, the economic dependence -in terms of funds- of the system from the central regulatory authority (Ministry of Education, University and Research - MIUR) remains valid.

Afterwards the European uniformity effort started with Bologna Process (1998) has driven the Italian system to a period of reform of the educational system. Bologna Process consisted of 46 volunteering countries that met with the aim to create a European Higher Education Area (EHEA). *"As the main objective* *of the Bologna Process since its inception in 1999, the EHEA was meant to ensure more comparable, compatible and coherent systems of higher education in Europe".* One of the major Italian effects due to this European Process was the reduction in time to completion.

According to the OECD report the Italian survival rate in 1998 was 35% with a drop-out rate equal to 65%, and a graduation rate of 9% with respect to the enrolled students. Finally also more than 80% of the system's economy was provided by the public sector. In response to these problems, a reform process started (University reform act, 1999) and the structure of the system was transformed from a unique course of four years to a two level educational system with the standard bachelor/master organization. Finally, PhD degrees represent the last educational step and would be more devoted to people that have an inclination towards joining academia, while their labour market value remains quite close to the master's (in some sense it should be also considerd lower thinking about the three years age difference of PhD students with respect to standard graduates). Hence in the last years the effort was both to coordinate the national system with the EHEA and to shorten the time to qualification. Then additional reform acts were not introduced till the last three years when a substantial reform process started with the advent of Berlusconi's last government (2008- November 2011).

The institution of an independent National Agency for the Evaluation of the University System and Research, (ANVUR) in 2006 (instead of the National Committee for the Evaluation of the University System – CNVSU), the introduction of an increasing share (starting from 7% in 2009) of the FFO divided between the university institutions according to a central evaluation exercise (Law n° 1, 2009) and the last Reform Act (14th January, 2011), which establishes, as a basic principle, a meritocratic criteria related to the efficacy and efficiency of the institutions in future evaluation gains. The guiding principles of this reform fix in quality, openness, efficacy, and efficiency the seeds/keywords of the future reorganization of the system.

All of these are signs of the probable new policy direction of Italian universities. We suppose that this changing process will be in the interest of the academic system in the upcoming years and the role of efficiency evaluation methods will be preeminent in the future setting. Actually the debate between experts and decision-makers about the choice of the specific evaluation criteria is a great issue in Italy, but indeed it lies outside our contingent purposes. The final aspect of

Italian system that we need to consider at this stage is that it is composed of 103 academic institutions grouped into 26 private universities, 61 public universities, ten distance universities and six high-quality education institutes.

The aims, purposes and juridical regime for all institutions are the same, what differs are nature and methods of their financing systems. Indeed the shares of funds collected from the private sector and the public sector are different between public and private academic institutions. Private funding mostly finances the budget of private universities. A minor share of their total budget is due to the public education system in a sort of private bargaining mechanism between each institution and MIUR. On the other hand the higher education institutes are autonomous institutions that provide especially higher quality and research related academic degrees (usually postgraduate and post-doctoral).

Finally, we note that the recent reforming process will depend on the concrete effort of the academic actors deputed to the reorganisation strategy. However, considering the international trend the evaluation exercise will have, starting in the short term and maybe realizing in the long run, an increasing importance in the national academic debate and in future government policies.

2.3 Data

The data is collected for inputs and outputs of universities in Italy for four academic years between 2005/2006 and 2008/2009 form the basis of our analysis. The integrated data-set is created combining information's from different sources.

Primary source is represented by MIUR that shares a wide variety of data on its public website. The university's information consists of the number of students enrolled, number of graduated students (both undergraduate and postgraduate togheter), total number of teaching staff and total number of academic staff per year in both private and public universities. Also, data on student facilities and accommodation services are provided (DSU).

Second we collect data on research outputs on ISI – Thomson web version dataset. This dataset is a collection of all the research products (papers, working papers, letters, books, conference proceedings) published in the last 40 years by more than 8.700 scientific journals. All these journals are published in English and have an anonymous referee evaluation system for each paper. Anonymity of the refereeing process usually implies stringent selection criteria of scientific quality, reserving the pages available (that are always very limited in the most important and prestigious journals) to the best of the scientific works submitted for publication. We know that this dataset is affected by some shortcomings. The most common critics of ISI citation database notes that it covers mainly North American, Western European (Italy), and English-language papers, it is limited to citations from 10.000 journals and provides different coverage between research fields. We are conscious of these potential problems and of the fact that this database probably is not the most representative for Italian Universityies because it could be biased mainly by differences in research fields that are own of each academic institution in the country. By the way it is the best available, international, quality-certified (using a worldwide standard quality measure: the 2010 Journal Citation Reports® -JCR-) source of published products and it consists of more than 250.000 papers, books, and conference proceedings referring to our 59 referring Universities in 2005/06-2008/09 academic years (more than 60.000 research products by year). Distance Universities, most of the small and recent private universities and three of the standard ones (L'Aquila, Arcavacata di Rende and Vercelli and) are excluded by the study due to data avilability. Measures of ISI products and teaching and research inputs and outputs were not available for them.

Figure 1 shows quantity (P) and quality (C and C/P) of ISI products by country in the eightees, what is important to note here is that Italy was the forth country for quantity of published works and the nineth for quality of scientific production (C/P). So ISI database could be considered an appropriate, standard and accepted source for scientific production information since 30 years and Italian Scientific Output is well-represented by this source (except for subject mix bias).

Nation	Impact (C/P)	Citations (C)	Publications (P)
Sweden	13.63	2 733 844	200 519
Denmark	12.80	1 277 433	99 836
Netherlands	12.36	3 029 184	245 055
UK	12.08	12 035 155	995 997
Belgium	10.43	1 273 803	122 106
Finland	10.31	875 172	84 881
France	9.73	6 200 504	637 357
Germany	9.61	8 391 278	873 536
Italy	8.76	3 070 391	350 433
Austria	8.51	666 197	78 266
Ireland	7.20	213 072	29 575
Spain	6.10	1 255 011	205 742
Portugal	5.59	111 770	19 979
Greece	5.20	231 153	44 466

Figure 1: Overall scientific production in EU countries, 1981-1990[2].

The integrated data set obtained yields information about all (business administration, teaching and research) inputs and outputs we consider in our production function.

Input includes both the absolute value of enrolled students and the number of enrolled students weighted by a measure of their skill-level at the end of the high-school (to capture both the quantity and quality of graduate teaching input). These two input measures will be used alternatively in the production function and joined with the graduate quantity-quality measure will show us the rising of a grade-inflation effect. Also measures of total number of post-graduates students enrolled (PhD_ENR) is included as a measure of post-graduate teaching and research input. The values reported for students and staff are not based on the standard convention of Full Time Equivalents (FTE) but they are the absolute values of these data. Unlikely Italian data do not allow a distinction between full-time and part-time students, so the FTE information is unavailable. Same information's are unavailable for teaching, research and administrative staff. The number of teaching and research staff are considered according to their academic position (Lecturers, Associate professors, Full professors) as a measure of research and teaching inputs. These measures reflects the structure of the academic staff without taking into account the permanent or temporary

[2] Breno, E., et al., (2002) "Scientific Research in Italian Univesities: an initial analysis of the citations in the ISI data bank", CRUI

status of their positions. We decided to consider all the individuals that take part into education and research process in the academy in the considered year. All of them contribute to realize the final process.

Other measures of research effort are taken by the ISI dataset, as the total number of citations by university and the total number of papers published on ISI journals and are considered as proper outputs. A measure of the quality of these papers is also included. It represents the number of paper published by each university weighted by the quality of the journal on which it is published .

Journal quality is given by the value of its 2005-2010 impact factor (2010 Journal Citation Reports®). In a nutshell we can summarize the Impact Factor meaning as a measure of the average number of citations expected by an article published in a certain journal (according to the definition of impact factor value). Outputs also include number of graduates weighted by their degree classification to capture both quantity and quality dimensions of teaching output. Finally Ph.D number of graduates (PhD) are included as output of the post-graduate teaching effort.

Table 1 reports descriptive statistics for the selected inputs and outputs. Total teaching staff (TEACH) is considered as an aggregate measure of all the academic positions. On average the academic population is composed by 960 professors, relatively higher is the number of lecturer (376) while Associate and Full positions are -on average- composed by 287 and 294 units. Standard deviation is very high because of great volatility between different universities. A similar number of staff (841) represents the average of non-academic workers.

Variable	Mean	Std.Dev.	Min	Max	Cases	Missing
Inputs						
TEACH	960,4	883,4	32	4736,0	236	0
FULL. PROF.	294,29	273,8	8	1.473	236	0
ASS. PROF.	287,9	253,4	6	1.370	236	0
LECTURER	375,9	356,5	7	2.025	236	0
STAFF	841,8	822,8	0	4.841	236	0
PhD. ENR.	199,56	197,7	0	1.293	236	0
PROJ.FIN.	14,0	16,3	0	104	236	0
DSU	3.006,04	2.745,9	0	16.823	236	0
TOTQ	123.723	106.416	2540	541.690	236	0
ENROLLED	4812.31	4342.29	35	24.310	236	0
Outputs						
GRADQUAL.TOT	95.358	81.421	1435	438.200	236	0
GRADUATES	4414.39	3953.97	60	20648	236	0
RESEARCH	2.883,67	3495,2	0.603	16.509,5	236	0
TOT. PAPERS	711,5	788,55	0	4.108,0	236	0
PhD	163,7	188,8	0	1.851	236	0

Table 1: Descriptive Statistics

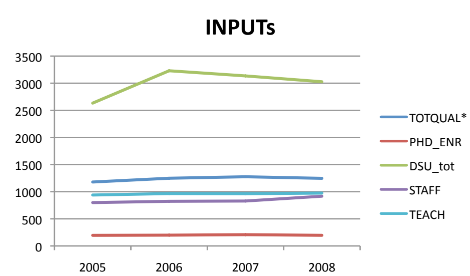


Figure 2: yearly level of Outputs and Inputs; *1% level of the original variable

Total number of enrolled students weighted by the initial skill level, absolute value of enrolled students (4.812), number of research project financed by the national research programme (14) and number of student's services provided by the universities (3.006) complete the whole set of inputs considered. Finally we consider the weighted number of enrolled students of five-academic years earlier as the input related to the current year graduates weighted output. Five years in Italy is the legal time for graduation and we suppose that it is the best proxy.

In the outputs box we can find the quantity-quality measure of teaching effort (Gradqual) and the only quantity measure (4.414), number of PhD students that obtained their degree and the direct measures of research weighted by the impact factor value (2.833).

Total n° of Paper by University				
Year	**Min**	**Max**	**Mean**	**Std.Dev**
2005	1	3.692	639.7	799,8
2006	1	3.931	689.5	849,9
2007	1	4.294	783.8	946,9
2008	4	4.414	817.2	965,5
Paper's citations (avg) by University				
Year	**Min**	**Max**	**Mean**	**Std.Dev**
2005	1	18,18	10.53	4,46
2006	1	32,56	9.01	4,80
2007	1	26,00	6.90	3,33
2008	1	9,22	4.75	1,89
Total sum of citations by University				
Year	**Min**	**Max**	**Mean**	**Std.Dev**
2005	1	45.772	7.562,1	10.186.6
2006	1	38.352	6.684,8	8903.1
2007	1	34.178	5.705,1	7587.1
2008	5	25.060	4.350,4	5807.9

Table 2: Descriptive Statistics of ISI informations

The measure of ISI research effort is available from 2001 to 2010 and it is composed by 450.845 papers written by at least one of the authors who is belonging to an Italian University. Then the ISI output is aggregated by university taking into account the average number of papers, the average number of citations and the total sum of citations by university and by year (Table 2), while the quality dimension of the research output is given by the impact factor value of the journals where papers were published. The impact factor measure is available online, thanks to ISI - Journal Citation Reports database, as the average impact factor of the journals over a five years period: 2005-2010. This quality measure is by definition fixed during the time span for the 10.477 journals considered in our analysis (2.731 journals in social sciences and 7.746 in science fields). Table 3 shows the average value of the impact factor and the average number of citations of the overall papers over period 2005-2008 [Table 3 and Figures 3(a) and (b)].

Year	N° Universities	avg if 2005/2010	avg tot_quot
2005	59	3.40	10,53
2006	59	3.41	9,01
2007	59	3.24	6,90
2008	59	3.30	4,75

Table 3: average impact factor 2005/2010 by year

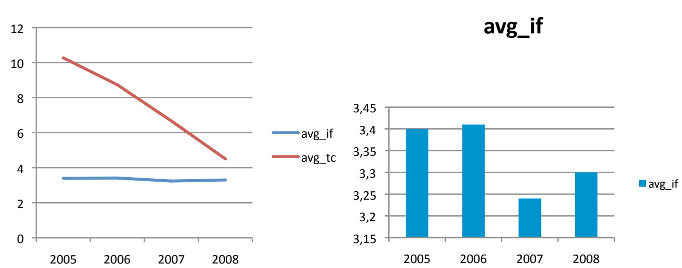


Figure 3: average impact factor and average number of citations by year (a); absolute value of the impact factor by year (b), 2005-2008.

The year by year measure is determined over 257.079 papers published in the period 2005-2008, a share of 77,4% of the ISI papers have been reconducted to their impact factor value (Figure 4). The average number of citations is influenced by time, therefore there is an expected positive relationship between time since publication (spread of knowledge amongst researchers) and number of citations for each paper. Hence the negative slope of the average number of citations is not a surprise and considering it as a measure of research quality could be misleading (Figure 3, a).

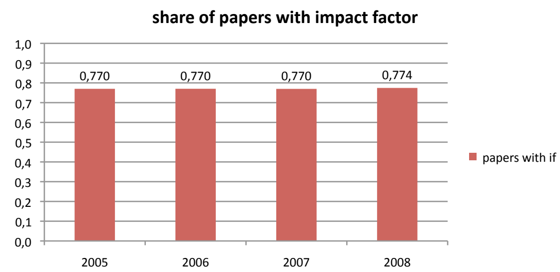


Figure 4: shares of ISI papers with impact factor (2005-2008)

Therefore we decide to use as our research quality measure the year-by-year quantity/quality score given by the weighted average of the number of paper and the 2005/2010 impact factor value of their publishing journals. Figure 5 shows the dynamics over time of research output versus the total number of papers by year. It seems that the quantity-quality measure of research effort is increasing in the time span.

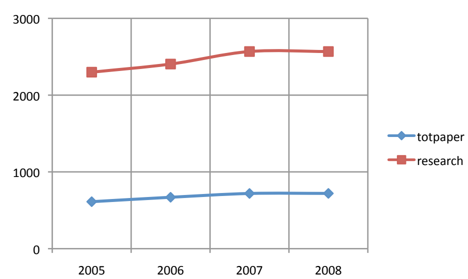


Figure 5: research output and total number of papers by year.

2.4 Methodology

Stochastic frontier models have become popular tools in productivity analysis. One of the interesting advantages of these specific models is the ability to account simultaneously the production relationship between inputs and outputs and the determinants of the inefficiency of institutions. Production function models can be estimated using both parametric and non-parametric methods. Both of these strategies are well-known in efficiency literature. It is reassuring that comparative studies of efficiency scores with both parametric and non-parametric tools show that these measures are highly correlated (correlation close to 0.70) (Ferrier and Lovell, 1990; Coelli and Perelman, 1999; Whiteman, 1999, Fecher et al., 1993). The most common non-parametric method is Data Envelopment Analysis (DEA). DEA is a linear programming method which examines the relationship between inputs to a production process (resources used in an institution) and the outputs of that process. It assumes that there is a frontier technology that can be described by a piece-wise linear hull that envelopes the observed outcomes. It is useful because it does not required a definition of the production function functional form for the technology, by the way serious drawbacks consist on the exclusion of stochastic error terms (that implies no measurement error or random fluctuations), strong sensitivity to outliers and obviously (by definition) no estimates of the production function parameters. All these shortcomings make the difference in our choice towards parametric methods.

Stochastic Frontier Analysis is an econometric technique which uses regression analysis to estimate a production function, with the difference being that the efficiency of an institution is measured using the residuals of the estimated equation where the error term is divided into a stochastic error term and a systematic inefficiency term. Stochastic frontier models date back to the cornerstone papers of Aigner, Lovell and Schmidt (1977) and Meesen and van den Broek (1977). They proposed a production function with a two-part composite error term in a cross-sectional setting as follows

$$y_i = \alpha + \beta' x_i + e_i \tag{1}$$

Where $e_i = v_i - u_i$ and v_i is a symmetric random error to account for statistical noise that arises from the inadvertent omission of relevant variables from the vector x_i, as well as from measurement errors and approximation errors associated with the choice of the functional form. Withal u_i is a one-sided random variable representing 'technical inefficiency' (Farrell, 1957), that is defined as the distance of the specific observation from the production estimated frontier (Figure 6). This could -by theory- follow any non-normal distribution, so that it can be separated out from the other residual term. A common assumption in literature is that it follows a truncated half-normal distribution. And we agree with this assumption in our article. The random error term (v_i) can be positive or negative and so the stochastic frontier outputs vary about the deterministic part of the model.

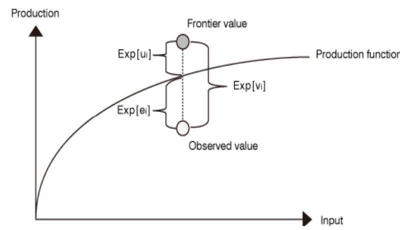


Figure 6: deviation from the production frontier obtained with SFA in a cross-section analysis[3]

In a standard set-up maximum likelihood methods are used to estimate regressor coefficients. Hence the most-common measure of technical efficiency is the ratio of the observed output to the corresponding stochastic frontier output

$$TE_i = \frac{q_i}{e^{\left(x_i'\beta + v_i\right)}} = \frac{e^{\left(x_i'\beta + v_i - u_i\right)}}{e^{\left(x_i'\beta + v_i\right)}} = e^{(-u_i)} \tag{2}$$

This measure of technical efficiency takes a value between zero and one. It represents the output of the $i - th$ institution relative to the output that could be produced by a fully-efficient firm using the same input vector. This notion of technical efficiency shows the ability of a firm, institution, country or university to obtain maximal level of output from a given set of inputs. It's measured by the output of the firm relative to that which it could attain if it was 100% efficient, or in other words if it lays on the frontier itself. Clearly the first step in predicting the technical efficiency, TE_i, is to estimate the parameters of the stochastic production function model. The production functions express a single dependent variable (single or aggregate ouptut measure) as a function of one or more explanatory variables (usually inputs).

There are various ways in which the relationship between inputs and outputs of higher education institutions can be explored. There could be two types of approach, an output oriented strategy where we assume that the instution aim is to maximize output from a given level of inputs and an input-oriented perspective where universities assume to minimize inputs from a fixed level of outputs. In studies where comporative analysis are estimated between these two choices results shown that the efficiency scores are strongly correlated (Coelli and Perelman, 1999). In the case of higher education we consider an output-oriented perspective as the more realistic strategy where universities try to maximize their

[3]Source: Japanese Ministry of Internal Affairs and Communications, Vol. 20, 2009

outputs (quantity or quantity-quality of graduates, number and quality of papers produced) from their own given level of inputs (academic staff, amount of public funds and/or grants, quantity or quantity-quality of enrolled students,...).

In this framework the efficinecy of an institution can be assesessed by studying outputs produced in relation to either the cost incurred or the inputs used. The former (cost) approach is the most common tool of analysis applied by researchers according to the literature (Cohn et al., 1989; Johnes, 1996; Hashimoto and Cohn, 1997; Glass et al., 1998; Izadi et al., 2002; Cohn and Cooper, 2004; Stevens, 2005; Worthington and Higgs, 2008; Johnes and Johnes, 2006). The question below this former approach is: by how much can cost be reduced without changing the output quantities produced by the institution? Therefore the cost functions can not be used to estimate the characteristics of multiple-output production technologies in cases where we have no price information (Perelman, Santìn, 2005) and/or it is inappropriate to assume firms minimise costs or maximise revenues (e.g. when industry is outside regulated). The HEIs sector lays exactly in this case, where institutions are settled in a non-profit sector where the problem is not to maximize revenues or minimize costs but to obtain some different and selected outputs starting from their level of inputs. So the study of the production of outputs referring to the inputs is the best strategy in the HE case. This approach is more complicated and there is only one paper, according to my knowledge, about the production relationship in the UK university sector (Johnes, 2010), and no studies, to date, on the Italian HE sector.

In a multiple-output/multiple-input context the simple alternative is to estimate the relationship of an aggregate output measure and the multiple inputs. But of course the aggregate composite output measure has to be defined by the researcher itself, estimating (that means arbitrarily deciding) some weights to assign over each output dimension. Consequently an estimation bias could affect this approach beacause of the incorrect (even if unintentionally) definition of the weights.

The distance function strategy offers an alternative which is not beset by any of these problems. Distance functions allow for both multiple inputs and multiple outputs (Coelli and Perelman, 2000; Rodriguez-Alvarez et al, 2004; Tonini, 2004), does not assume any particular optimizing behaviour on the part of the institutions, does not require a knowledge of prices of either inputs or outputs (Coelli and Perelman, 1999; Coelli, 2000; O'Donnel and Coelli, 2003; Uri, 2003) and does not require prices to be exogenous (Banos-Pino et al, 2002). All of these points make the distance function approach particularly attractive in the context of higher education (Johnes, 2010).

2.4.1 Estimating a university production function

In this paper we propose a parametric stochastic distance function at instituion level in order to go beyond in the efficiency analysis of higher education sectors. For this purpose our function could be represented by the following

$$D_{it} = f\left(R_{it}, T_{it}, M_{it}\right) \tag{3}$$

Where f represents the best technology used in transforming institutional inputs in outputs, D_{it} is the distance that separates each university (i) from the technological boundary at time t, R_{it}, T_{it}, and M_{it} represent Research, Teaching and Managerial dimensions (measured through their inputs and outputs) of each educational institution (i) at time t. This function is settled naturally in a parametric stochastic frontier analysis (SFA) framework.

The SFA distance function

First of all we have to define a vector of inputs $x = (x_1, ..., x_k) \in \mathbb{R}^{k^+}$ and a vector of outputs $y = (y_1, ..., y_z) \in \mathbb{R}^{z^+}$, so that the multi-input/multi-output production technology can be defined as the output distance function (Shephard, 1970): $D_{it}\left(x, y\right) = inf\left(\theta : \theta > 0, \left(x, y/\theta\right) \in P\left(x\right)\right)$ where $P\left(x\right)$ is defined as the output possibility set that can be produced using the input vector x. If $D\left(x, y\right) \leq 1$ then the institution possibility set $\left(x_{it}, y_{it}\right)$ belongs to the production possibility set $P\left(x\right)$, otherwise if $D_{it}\left(x, y\right) = 1$ it means that y is located on the outer boundary of the output possibility set.

In order to estimate in our context the defined distance function in a parametric framework a trans-log functional form is proposed. Coelli and Perelman (1999) demonstrate that this trans-log specification fulfills a set of desirable characteristics: it's flexible, easy to derive and allows the imposition of homogeneity of degreee one with little trouble. The trans-log has also been use to estimate distance functions in many different context: telecommunications (Whiteman, 1999; Uri, 2003); railways (Coelli and Perelman, 1999; Banos et Pino, 2002; Atkinson et al. 2003; O'Donnel and Coelli, 2005); electric utilities (Whiteman, 1999; Atkinson and Primont, 2002; Atkinson et al., 2003); water industry (Saal and Parker, 2006); and agriculture (Paul et al, 2000; 2002; Karagiannis et al., 2004; Tonini, 2004; Paul and Nehring, 2005; Balcombe et al., 2007). Translog distance function is defined for the HEIs using inputs $x_k\,(k = 1, ..., K)$ to produce outputs $y_z\,(z = 1, ..., Z)$:

$$lnD_{it}\,(x, y) = \alpha_0 + \sum_{z=1}^{Z} \alpha_k lny_{zit} + \frac{1}{2} \sum_{z=1}^{Z} \sum_{h=1}^{Z} \alpha_{zh} lny_{zit} lny_{hit} + \sum_{k=1}^{K} \beta_k lnx_{kit} +$$
$$\frac{1}{2} \sum_{k=1}^{K} \sum_{j=1}^{K} \beta_{kj} lnx_{kit} lnx_{jit} + \sum_{k=1}^{K} \sum_{z=1}^{Z} \gamma_{kj} lnx_{kit} lny_{zit}$$

where i denotes the $i-th$ university at time t. In order to obtain the production frontier surface we set $D_{it}\,(x, y) = 1$ which means that $lnD_{it}\,(x, y) = 0$. According to most of the literature a trans-log distance function allows the possibility of applying linear regression techniques -thanks to its logarithmic form- and the satisfaction of most of the subsequent criteria (Coelli, Rao, Battese, 1999): they have to be linear in inputs and outputs, non-decreasing, linearly homogeneous and concave in outputs. Trans-log distance functions have the characteristic of being a satisfactory functional form for the first two assumptions and to the third one applying simple constraints to obtain the homogeneity of degree + 1 in outputs. According with Lovell et al. (1994), to normalize the output distance function by one of the outputs is equivalent to imposing homogeneity of degree + 1.

Homogeneity of degree +1 in outputs conditions:

$$\sum_{z=1}^{Z} \alpha_z = 1$$

$$\sum_{h=1}^{Z} \alpha_{kh} = 0, z = 1, 2, ..., Z$$

$$\sum_{k=1}^{Z} \gamma_{kh} = 0, k = 1, 2, ..., K$$

Symmetry conditions:

$\alpha_{kh} = \alpha_{hk}, k, h = 1, 2, ..., Z$

$\beta_{kj} = \beta_{jk}, k, j = 1, 2, ..., K$

By the homogeneity condition $D(x, \rho y) = \rho D(x, y)$ which means that the output can be chosen arbitrarily, for axample the Z-th output, such that $\rho = 1/y_Z$ and rearranging the function this way we could obtain the following equation:

$$-ln(y_{Zit}) = TL(x_{it}, y_{it}/y_{Zit}, \alpha, \beta, \gamma) - lnD_{it}(x, y) \quad (4)$$

where $-lnD_{it(x,y)}$ corresponds to the radial distance function from the boundary. According to Battese and Coelli (1988) specification we obtain the distance function version of SFA setting $u_{it} = -lnD_{it}(x, y)$ and adding up a term v_{it} capturing the noise

$$-ln(y_{Zit}) = TL(x_{it}, y_{it}/y_{Zit}, \alpha, \beta, \gamma) + \epsilon_{it}, \epsilon_{it} = v_{it} + u_{it} \tag{5}$$

where $u_{it} = -lnD_{it}(x, y)$ represents the distance to the boundary set, and is a negative random term assumed to be independently distributed as truncations at zero of the $N(0, \sigma_u^2)$ distribution, and the v_{it} term is assumed to be a two-sided random (stochastic) disturbance designated to account for statistical noise and distributed $iidN(0, \sigma_u^2)$. The truncated-normal distribution of the u-term allows us to relax the hypothesis that the mean of the underlying inefficiency variable is zero (as in the normal-half normal case), and this seems to be the most realistic assumption. Hence v_{it} and u_{it} are independent of each other, independently and identically distributed such that $v_{it} \sim N(0, \sigma_v^2)$ and $u_{it} \sim N^+(0, \sigma_u^2)$ and N^+represents a truncated-normal distribution truncated at 0. Thus $-u_{it} = -lnD_{it}(x, y)$. The decision about the method of estimation depends on which is the assumption we make on the technical inefficiency component u_{it} which contains components related both to institutions and time. The simplest assumption is the one that allows inefficiency to vary by institutions and not by time (time-invariant; $u_{it} = u_i$). In this simple case u_i can be treated as a fixed parameter (Coelli et al., 2005). A second alternative is to assume a time-varying decay model for the inefficiency term (as proposed by Battese and Coelli, 1995). This way $u_{it} = \{exp[-\eta(t - T_i)]\}\, u_i$ where $T_i$is the last period in the *i-th* panel, η is a decay parameter to be estimated by the model and u_i is the base-level of inefficiency which is the inefficiency in the last period observed for unit i. Inefficiency decreases towards the final period level (over time) if $\eta > 0$ and vice-versa an estimated value of $\eta < 0$ means that inefficiency increases towards the final period. To our knowledge a more appropriate estimation strategy of the inefficiency term with respect to the previous ones is the random parameters approach (Greene, 2005; Tsionas, 2002). This approach allows the production function to vary more generally across institutions to account unmeasured heterogeneity as well as firm inefficiency within our statistical framework. The potential surplus regarding this approach is illustrated by Johnes (Johnes et al. 2008 and Johnes and Johnes, 2009) with a comparison study of two cost-functions over HEIs in England. A standard stochastic frontier production function and a random parameters version of the same cost-efficiency model are estimated using the same data-set over UK Universities and they found that mean effiiciency is increased of 7% (from 69% to 75%) allowing one of the outputs parameter to vary by Institutions.

Random parameters technology

In the present paper we estimate the model on a panel data set of four years, so (1) needs to be modified as follows

$$y_{it} = \alpha + \beta' x_{it} + v_{it} - u_{it} \tag{6}$$

so that y_{it} is the performance of institution i in period t (output), x_{it} is the vector of inputs and outputs, the β_i are modelled as random parameters and we assume that these follow a normal distribution. According to our trans-log distance function approach eq. (6) becomes the following (defining $\beta^* = (\alpha, \beta, \gamma)$ in the previous notation)

$$-ln(y_{Zit}) = TL(x_{it}, y_{it}/y_{Zit}, \beta^*) + v_{it} - u_{it} \tag{7}$$

Greene (2005), extending the works of Tsionas (2001), Kumbhakar (1990) and Battese and Coelli (1995) proposed this (6) random parameter version of (3) allowing the variance of the inefficiency term over time. For long time series data, time invariance is a too strong assumption in stochastic frontier models (Pitt and Lee, 1981). A general form of the random parameters stochastic frontier model may be written as

(1) Stochastic frontier:

$y_{it} = \alpha_i + \beta_i' x_{it} + v_{it} - u_{it}$,

that in our study is $-ln(y_{Zit}) = TL(x_{it}, y_{it}/y_{Zit}, \beta^*) + v_{it} - u_{it}$

$v_{it} \sim N[0, \sigma_v^2], v_{it} \perp u_{it}$

(2) Inefficiency distribution:

$u_{it} = |U_{it}|, U_{it} \sim N[\mu_i, \sigma_{ui}^2]$,

$\mu_i = \mu_i' z_i$,

$\sigma_{ui} = \sigma_u exp(\theta_i' h_i)$.

(3) Parameter Heterogeneity:

$(\alpha_i, \beta_i) = (\bar{\alpha}, \bar{\beta}) + \triangle_{\alpha,\beta} q_i + \Gamma_{\alpha,\beta} w_{\alpha,\beta_i}$,

$\mu_i = \bar{\mu} + \triangle_\mu q_i + \Gamma_\mu w_{\mu_i}$

$\theta_i = \bar{\theta} + \triangle_\theta q_i + \Gamma_\theta w_{\theta_i}$

Each subvector of the full parameter vector (α_i, β_i), μ_i or θ_i is allowed to vary randomly with means vector $(\bar{\alpha}, \bar{\beta}) + \triangle_{\alpha,\beta} q_i$ and likewise for the others, where $\triangle_j$is a standard matrix of parameters to be estimated and q_i is a set of related variables which enters the distribution of the random parameters.

The w_{ji} (where $j = (\alpha, \beta), \mu, \theta$) vector contains the parametrization of the random variation and it is assumed to have mean vector zero and known diagonal covariance matrix Σ_j. The random vectors w_{ji} is usually assumed to be normally distributed (and so $\Sigma_j = I$).

Therefore the log-density for the random parameters version for stochastic frontier model is $logL_{it} = logf(\Theta_i | x_{it}, z_i, h_i, q_i, w_i)$ where Θ_i contains all the parameters of the model, e.g., for (α_i, β_i), this is $(\bar{\alpha}, \bar{\beta}, \triangle_{\alpha,\beta}, \Gamma_{\alpha,\beta})$ and likewise for μ_i and θ_i. Then assuming that conditioned on the firm specific random parameters vector w_i the observations are independent, the conditional log-likelihood for the sample is $logL | w_1, ..., w_N = \sum_{i=1}^{N} \sum_{t=1}^{T} logf(\Theta_i | x_{it}, z_i, h_i, q_i, w_i)$. And in order to estimate the model parameters it is necessary to integrate the heterogeneity out of the log-likelihood. The unconditional log-likelihood is $logL = \sum_{i=1}^{N} \int_{w_i} \sum_{t=1}^{T} logf(\Theta_i | x_{it}, z_i, h_i, q_i, w_i) g(w_i) dw_i$ where $g(w_i)$ is the multivariate density of the random vector w_i.

However there is no closed form for density of the compound disturbance in this model, so integration can be done only by quadrature (for small models with less than two random parameters) or by simulation (MSL-Maximum Simulated Likelihood).

Estimation of the random parameters model is extremely time-consuming and the simulation algorithms can achieve a reasonable approximation to the true likelihood function for a large number of random draws. So the process of maximization can be done using 'intelligent' draws, such as Halton sequences (Bhat, 1999; Train, 2002), that can reduce the number of required draws by a factor of five or ten. To our porpouses we fitted the model using 1000 Halton draws, which is roughly equivalent to random simulations of several thousands of draws.

Accounting for Background variables

Question of how to model environmental (or background) variables in stochastic frontier analysis is seldom discussed in economic literature. Previous studies applied a two-step approach using these regressors as determinants of efficiency in a second stage regression of the predicted efficiency scores on university-specific factors. Others include them into the frontier itself. But the question in this case is: do these factors affect production directly or indirectly? (Stevens,P.A., M., Vecchi, 2005). In the first case the environmental factors enter the production frontier function, while in the latter case they affected indirectly through the effects on the efficiency. The big problem of the direct approach is due to the theoretical hypothesis that are implied: in the first frontier step the efficiency scores are assumed to be independently and identically distributed while in the second regression step we assume that they are a function of some university-specific factors, implying that they cannot be identically distributed (Kumbhakar, Gosh and McGulkin, 1991). A popular method that overcomes this problem is the one proposed by Battese and Coelli (1995), which consists in estimating both frontier and efficiency terms in a unique stage. Battese and Coelli extended the works of Reifschneider and Stevenson (1991) and Huang and Liu (1994) to the panel framework. For these reasons exogenous variables could be considered as background or environmental variables in that each university has little control over them, in the efficiency literature they are named as 'non-discretionary' variables (Worthington, 1999; Stevens and Vecchi, 2002). These factors affect the attractiveness of the academies on high school students, they give a different market power to each institution because of their subject-mix and others similar factors.

According to these reasons we introduce out from the trans-log distance function some additional variables to represent trend over time and differences between groups of universities. We fit our model incuding two dummy variables (state vs private and polytechnic vs others), a variable measuring the rate of non-regular students per institution ('perc') and a variable capturing subject-mix.

This subject-mix variable consists on the proportion of students enrolled in science and technical courses as a proxy of departmental composition of the university.

In order to attain a rigorous assessment of HEIs, we are conscious of the need to take into account differences in the subject-mix composition of institutions regarding the possible impact that these differences may have in explaining different performances. Since disciplines vary in terms of research productivity and research outputs, these differences need to be considered when we are measuring quantitative performances of universities, otherwise we may draw spurious conclusions regarding the relative performances of Italian academic institutions (Sarrico and Dyson, 2004). Probably the better solution to this problem in our particular context would be represented by a less aggregate efficiency analysis which measures efficiency at departmental or faculties level (Sarrico, Teixeira et al., 2009). But serious drawbacks on this solution consist in data availability, inputs and outputs measures at this lower level are unavailable for the Italian system. The purpose of the author is to insert a variable that has the role of measure the ISI exposure of each University. The higher it is this variable greater is the exposure in terms of subject-mix composition of the specific academic insititution towards disciplines devoted to ISI in terms of research outputs.

2.5 Model estimation

In order to investigate inefficiency of Italian HEIs, we employ a multidimensional production functions specifically without any direct influence of costs but only related to teaching, managerial and research efforts (a DEA approach of the same type is given by J. Johnes, 2006), whose basic set-up, according to our pourposes, is a trans-log function as specified earlier (7).

We considered different measures of output according to the three different dimensions of the HEIs: teaching effort, research effort, organizational issues.

Usually research inputs are considered in literature as proxies of research outputs, justified with the idea that *"research grants are in general awarded to meritorious groups of researchers on the basis of the quality and quantity of their previous work"* (Johnes, 1997, pag. 728). However some studies identified the correlation between research funding and quality research outputs (as the average score across all the departments in UK using the 1996 RAE data) in 0.65, and in 0.7 the correlation between faculty publications and grant support (based on data from National Academy of Sciences).

The estimation accuracy using these inputs instead of outputs is roughly low with respect to a direct measure of the quantity and quality of publications that has to be unavoidably better. For these reasons in this paper we use informations from the ISI-Thomson database. Research output measures of this source consist in the total number of papers published by each academic instititution in Italy weighted by the average impact factor (2005-2010) of the journal where they were published.

From the teaching point of view previous studies have rarely taken into account both the quality of the students enrolling and the quality of the final graduates. Few examples are Koshal and Koshal (1999), who used the average Student Aptitude Test Scores of input students, or Johnes and Johnes (2009). A failure to account for input and output quality would provide an imprecise measure of university teaching input and output. So it is clear that a correctly-specified model is important when making policy recommendations. We believe that the proportion of first and upper second class degrees represents a relatively consistent measure of Italian degree quality, and certainly the best that is readily available to our knowledge. Therefore as a first attempt the teaching outputs considered in this paper are the numbers of graduates (accounted for their quality through degree classification measures) and the number of postgraduate students. Stevens (2005) and Koshal and Koshal (1999) have found that the effect of the omission of output quality measures on the efficiency scores has ambiguous effects (in some case positive and in other negative). But for sure the omission of a measure of output quality implicitly assumes that all the HEIs are producing graduates of the same level of quality. And this is a really strong and roughly unbelievable assumption. We will see that the inclusion of quality dimension of teaching inputs and outputs in the Italian case has a misleading negative effect on efficiency score because of the grade-inflation effect. So both models with quantity-quality and only quantity measures will be fitted.

Disaggregation by subject areas (arts, sciences, humanities, engineering, ...) of the quantity (and quality) measures of graduates output (and input) were available in our information sources, but this choice has a multiplicative effect on the trans-log regressors number that consists on an unacceptable loss of degrees of freedom for our case. For this reason our production models in HE, as in the most common effiiciency literature, tend to be estimated (few exceptions to this rule are Johnes, 1997; Izadi et al., 2002; Stevens, 2005) using aggregate outputs (Athanassopulos and Shale, 1997; Johnes, 2008; Avkiran, 2001).

While the output measures for the research effort are the total number of paper published by each university weighted by their impact factor values (RESEARCH), the number of graduates (GRADUATED) and alternatively it's quantity-quality measure (GRADQUAL) and the number of students that obtained a PhD (PhD).

According to our first specification of the model the inputs variables are given by measures of research inputs, teaching inputs, management and services inputs. The total number of enrolled students (ENROLLED), also accounted for their quality using information of high school degree-levels (TOTQ), total number of post-graduate enrolled students (PhD_ENR), number of research project financed by the Italian Ministry of Education, University and Research as a research input measure (PROJ.FIN), total number of student's services and facilities (DSU), number of management workers (STAFF) and number of teaching staff (TEACH) are included.

2.6 Empirical Results

As usual in trans-log function estimations all the original inputs and outputs are size corrected and mean-corrected prior to estimation (Cuesta and Orea, 2002; Cuesta and Zofio, 2005). That means, each output and input variable is corrected by the size (given by the total number of enrolled students) and has also been divided by its geometric mean. In this way, the first-order coefficient can be interpreted as distance elasticities evaluated at the sample means.

In addition homogeneity of degree +1 is imposed by selecting one of the outputs, research in our study, as the dependent variable and the ratio of the other outputs with respect to it as explanatories in the trans-log function. First of all a simple random effects model of the first trans-log multi-input/multi-output production function was performed in order to underline that the output dependent composite measure has an effective strong relationship with our input regressors. The random effects results are reported in the table below, what we want to stress the skewness of the residuals, that is negative (-0.0809). A positive skewness is usually considered problematic in stochastic frontier studies because it cannot be reconciled with a one-sided distribution of inefficiencies that is positively skewed. Waldman (1982) proved that when Random Effects residuals are skewed to the left, then the Maximum Likelihood Estimator (MLE) for the frontier model is unique and we have no trouble in the estimation procedure of the frontier itself. Otherwise a *"positive skew implies that target institutions are 'super efficient' rather than inefficient"* (Green and Mayes, 1991, p. 528).

In table 4 we propose in addition to the non-frontier model the results of three stochastic frontier models, a pooled time invariant model, a time varying decay model (Battese and Coelli) and finally a fuller random parameters specification of the trans-log stochastic frontier production function with coefficients of inputs and outputs that are allowed to vary across institutions. In all cases the random parameters are assumed to follow a normal distribution. The model is solved by a simulated maximum likelihood technique using a specific econometric software (Limdep). We implement an Halton sequence of quasi-random draws to generate enough simulations that allow us to aproximate the real unconditional log-likelihood function and maximize the unclosed integral of this function (Tsionas, 2002; Greene, 2005). In fact it is not possible to solve the log-likelihood function with simpler techniques due to the existence of the unclosed integral.

		RE		TI SFA		TVD SFA (BC)		RPM SFA*	
Variable	**Param**	**Coeff**	**P-value**	**Coeff**	**P-value**	**Coeff**	**P-value**	**Coeff**	**P-value**
Constant		0,0868	0,4470	0,2101	0,0849	0,2187	0,0313	-0,4061	0,0000
Grad	α_1	0,2970	0,0000	0,3208	0,0000	0,2975	0,0000	0,2137	0,0000
PhD	α_2	-0,2145	0,0000	-0,2305	0,0000	-0,2114	0,0000	-0,1403	0,0000
Research	α_3	0,9175		0,9097		0,9131		0,9266	
Grad*Grad	α_{11}	-0,1813	0,0003	-0,1648	0,0000	-0,1725	0,0001	-0,0254	0,0002
PhD*PhD	α_{22}	-0,1851	0,0000	-0,1822	0,0000	-0,1825	0,0000	-0,0184	0,0018
Research*Research	α_{33}	0,325		0,2998		0,3106		0,025	
Grad*PhD	α_{12}	0,3457	0,0001	0,3234	0,0000	0,3328	0,0000	0,0344	0,0054
Grad*Research	α_{13}	-0,1644		-0,1586		-0,1603		-0,009	
PhD*Research	α_{23}	-0,1606		-0,1412		-0,1503		-0,016	
Teach	β_1	0,1721	0,0947	0,2027	0,0387	0,2097	0,0222	0,6678	0,0000
Staff	β_2	0,0666	0,1706	0,0808	0,1143	0,0752	0,0832	-0,0498	0,0000
Enrolled	β_3	0,3455	0,0001	0,2743	0,0004	0,2943	0,0000	0,14521	0,0000
Prog.fin	β_4	0,0546	0,0583	0,0597	0,0605	0,0574	0,0271	-0,04178	0,0000
PhD Enr	β_5	0,1708	0,0001	0,1658	0,0006	0,1597	0,0000	0,0774	0,0000
Dsu	β_6	0,0110	0,6156	0,0245	0,2421	0,0132	0,5039	-0,0195	0,0000
Teach*Teach	β_{11}	-0,1445	0,6549	-0,1147	0,7233	-0,1460	0,6177	0,0277	0,5706
Staff*Staff	β_{22}	0,0190	0,5466	0,0157	0,6861	0,0141	0,6170	-0,0206	0,0000
Enrolled*Enrolled	β_{33}	-0,2646	0,3541	-0,1487	0,6463	-0,1968	0,4394	0,2332	0,0000
Prog.fin*Prog.fin	β_{44}	-0,0823	0,0595	-0,0974	0,0615	-0,0874	0,0260	0,0087	0,2331
PhD Enr*PhD Enr	β_{55}	0,0356	0,0009	0,0351	0,0041	0,0336	0,0004	0,0156	0,0000
Dsu*Dsu	β_{66}	0,0007	0,8573	0,0047	0,2581	0,0008	0,8101	-0,0053	0,0000
Teach*Staff	β_{12}	-0,3391	0,1752	-0,2758	0,2793	-0,2671	0,2247	-0,0328	0,4730
Teach*Enrolled	β_{13}	0,1165	0,7732	0,0546	0,9032	0,2220	0,5374	0,1078	0,2644
Teach*Prog.fin	β_{14}	-0,0242	0,8976	-0,0357	0,8423	-0,0382	0,8200	0,1010	0,0018
Teach*PhD Enr	β_{15}	0,1695	0,2244	0,1712	0,3305	0,1839	0,1468	-0,0772	0,0000
Teach*Dsu	β_{16}	0,0282	0,6351	-0,0030	0,9676	0,0349	0,5161	0,0240	0,0203
Staff*Enrolled	β_{23}	0,6238	0,0047	0,6226	0,0051	0,5958	0,0023	0,1263	0,0006
Staff*Prog.fin	β_{24}	0,3008	0,0004	0,3066	0,0001	0,3015	0,0001	-0,0391	0,0054
Staff*PhD Enr	β_{25}	0,0592	0,4315	0,0604	0,5642	0,0524	0,4357	0,1281	0,0000
Staff*Dsu	β_{26}	-0,0943	0,1310	-0,1230	0,0334	-0,1175	0,0332	0,0325	0,0001
Enrolled*Prog.fin	β_{34}	-0,2156	0,2102	-0,2716	0,1082	-0,2545	0,0955	-0,1744	0,0000
Enrolled*PhD Enr	β_{35}	0,3690	0,0631	0,3842	0,0907	0,3603	0,0427	0,2343	0,0000
Enrolled*Dsu	β_{36}	-0,0408	0,4201	0,0107	0,8881	-0,0441	0,3402	0,0232	0,0033
Prog.fin*PhD Enr	β_{45}	-0,0063	0,8943	-0,0027	0,9609	-0,0065	0,8788	0,0032	0,7501
Prog.fin*Dsu	β_{46}	-0,0036	0,8584	-0,0054	0,8812	-0,0039	0,8337	-0,0126	0,0034
PhD Enr*Dsu	β_{56}	-0,0135	0,4098	-0,0055	0,8957	-0,0110	0,4501	-0,0171	0,0000

		RE		TI SFA		TVD SFA (BC)		RPM SFA*	
Variable	**Param**	**Coeff**	**P-value**	**Coeff**	**P-value**	**Coeff**	**P-value**	**Coeff**	**P-value**
Teach*Grad	δ_{11}	-0,4926	0,0221	-0,4342	0,0084	-0,4758	0,0140	-0,0627	0,0279
Teach*PhD	δ_{12}	0,4382	0,0297	0,3939	0,0099	0,4246	0,0191	0,0403	0,1443
Teach*Research	δ_{13}	0,054		0,0403		0,0512		0,0224	
Staff*Grad	δ_{21}	0,0078	0,9324	0,0311	0,6930	0,0316	0,7000	0,0053	0,7266
Staff*PhD	δ_{22}	-0,0007	0,9936	-0,0214	0,7870	-0,0135	0,8640	-0,0379	0,0059
Staff*Research	δ_{23}	-0,0071		-0,0097		-0,0181		0,0326	
Enrolled*Grad	δ_{31}	0,2064	0,3040	0,2319	0,2741	0,2480	0,1700	0,2945	0,0000
Enrolled*PhD	δ_{32}	-0,1094	0,5985	-0,1419	0,5342	-0,1469	0,4345	-0,3015	0,0000
Enrolled*Research	δ_{33}	-0,097		-0,09		-0,1011		0,007	
Prog.fin*Grad	δ_{41}	0,0239	0,7378	0,0229	0,7409	0,0233	0,7153	0,0909	0,0000
Prog.fin*PhD	δ_{42}	-0,0237	0,7238	-0,0271	0,7220	-0,0267	0,6584	-0,0696	0,0000
Prog.fin*Research	δ_{43}	-0,0002		0,0042		0,0034		-0,0213	
PhD Enr*Grad	δ_{51}	0,0680	0,2088	0,0633	0,2416	0,0657	0,1785	0,0149	0,1165
PhD Enr*PhD	δ_{52}	-0,0130	0,7656	-0,0101	0,8116	-0,0123	0,7544	0,0160	0,1171
PhD Enr*Research	δ_{53}	-0,055		-0,0532		-0,0534		-0,0309	
Dsu*Grad	δ_{61}	-0,0421	0,0623	-0,0550	0,0377	-0,0427	0,0323	0,0027	0,6309
Dsu*PhD	δ_{62}	0,0353	0,1210	0,0461	0,0915	0,0371	0,0659	0,0097	0,0781
Dsu*Research	δ_{63}	0,0068		0,0089		0,0056		-0,0124	
Time	τ	0,0021	0,3549	0,0020	0,0285	0,0021	0,0179	0,0058	0,0000
Health		0,0318	0,0591	0,0284	0,3185	0,0157	0,5980	-0,0517	0,0000
SciTech		0,0557	0,0031	,151301D-06	0,9996	,149153D-06	0,9980	0,0277	0,0011
Perc		-0,6107	0,2205	-0,6469	0,002	-0,6975	0,0001	-0,2868	0,0000
Type of University		0,1049	0,0267	0,0625	0,4243	0,0700	0,3572	0,4963	0,0000

Random Parameters SD of (**)	**RPM SFA***	
Constant	0,1811	0,0000
Grad	0,0176	0,0000
PhD	0,0406	0,0000
Teach	0,1140	0,0000
Staff	0,0029	0,1749
Enrolled	0,0943	0,0000
Prog.fin	0,0580	0,0000
PhD Enr	0,4426*E-04	0,9479
Dsu	0,0019	0,0000

	RE			**TI SFA**		**TVD SFA (BC)**		**RPM SFA***	
	Param	Coeff	P-value	Coeff	P-value	Coeff	P-value	Coeff	P-value
Sigma	σ^2			0.779	0.000	0.2171	0.001	0.1600	0.000
	σ_u^2			0.5956	0.000	0.0443	0,000	0.00411	0,000
	σ_v^2			0.0113	0,000	0.0028	0,000	0.00001	0,003
	η					0,0508	0,0809		
Log-Likelihood				152.86		248.15		291.308	
N		236		236		236		236	
N° of groups				59		59		59	
Technical Efficiency				0.9189		0.893		0.9378	
R-square		0.99							
Skewness (residuals)		-0.0809							
Kurtosis (residuals)		3.1183							

Table 4 - Parameter estimates of OLS and translog output distance functions (Time Invariant Random Effects, Time Varying Decay model and Random Parameters); Estimated parameters without p-values are calculated using the homogeneity conditions (see paragraph 0.01); * Coefficients reported for inputs and outputs is the mean; ** estimates of SD of random parameters normal distributions.

The estimated parameters of the models are presented in table 4. The time invariant Battese and Coelli specification of the model shows contraddictory results. Most of the coefficients are non significant, the Eta parameter of time varying inefficiency is positive, which should mean that the level of inefficiency is decreasing towards the final period level (considered the base level) and so over time. But the significance of technical progress in time variant model is neglected using the test of $H_0 : \eta = 0$ which means no certain evidence of progress in techinical efficiency over time.

Kernel density estimators are used in this case to compare the efficiency estimates from the pooled time invariant model to the Battese and Coelli model. Results are surprising, and not encouraging. The estimates of u_i from time varying Battese and Coelli model are far larger than those from the pooled model, which ignores the panel element of the data set (Figure 7, graph on the right).

Usually an important benefit of Battese and Coelli time varying model is that when data are very consistent with the model, it produces quite satisfactory results. But, a serious drawback is that, when this is the wrong model, extreme results can emerge (Greene, 2004). In our case there is a poor fit to this model, and results demonstrate this. Probably this is due to the assumption of time invariance in u_i that severely distorce the estimated model and it's inefficiency estimates.

However the presence of technical inefficiency in the time invariant model is assessed using the test $H_0 : \sigma_u^2 = 0$ against the one-sided alternative, rejecting H_0. So inefficiency is significant. Mean efficiency over the whole period is around 93% and varies from around 61,9% in the worst-performing university to 99% at the top university (on the basis of the random parameters sfa results; TI model suggests a minimum of 60% and a maximum score of 100%). We will explore later on the characteristics of the best-and worst-performing institutions. In the context of previuos results about mean efficiency in the Italian university system the only study, to my own knowledge, is Agasisti and Johnes (2010) paper on cost efficiency analysis over the period 2001-2003. Mean efficiency estimated by them was 81%. It is difficult to make comparisons between this study and the Agasisti-Johnes cost-efficiency analysis because of differences in the methodological approach (cost with respect to production perspective) and sample data (our data refers to 2005-2008 years and does not include any expenditure-related variable). Also efficiency rankings obtained are really different in the two studies. Referring to previous studies in other countries we can note that the range for production efficiency in English Universities is from 85% to 95% (Johnes, 2008; Flegg et al. 2004; Flegg and Allen, 2007) but it should, however, be borne in mind that efficiencies in each country study are defined in relation to the country-specific frontier, so it is hard to make comparisons between our estimation and other non-italian studies.

Patterns of efficiency over time are indicated by the coefficient of the time variable (T). Evaluated at the mean of inputs and outputs, technology has increased, in fact the frontier has shifting outwards by 0.2% to 0.58% (by year). Figure 7 plots also the distributions of the estimates of inefficiencies from the random parameters model (graph on the left) and the simple stochastic random effects frontier time invariant model (middle figure). The figures suggest that Random Parameters formulation is moving some of the variation of the production distance function out of the inefficiency term and into the production model, in the form of parameter estimates variation. It also appears that the random specification is moving the inefficiency to the left and reducing the variation. Which means that both mean and standard deviation are generally smaller than for the simpler, homogeneous parameters model. These results are statistically significant (p-values are lower than 0.05) and imply the existence of a positive technology change in Italian higher education system over the considered period. In general our random parameters results are plausible and suggest that this approach to frontier estimation can be extremely instructive in identifying inter-institutional differences in efficiency.

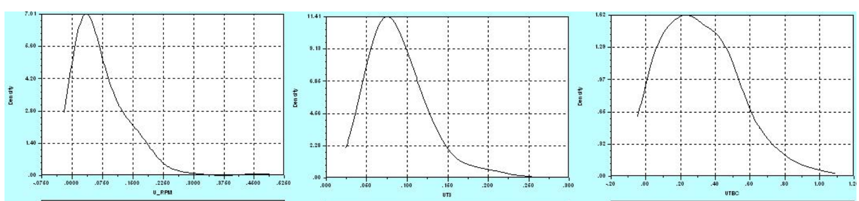


Figure 7 - Kernel density estimator for Random Parameters, Time Invariant and Battese andCoelli models Inefficiencies.

2.6.1 Efficiency scores

It's useful to see the characteristics of high and low-performing HEIs (Figure 8 a) and b)). Those HEIs in the highest efficiency quartile have the largest number, on average, of graduates-by size and PhD obtained by size (teaching and teaching/research outputs) while the level of research is on higher than the average, but slighty lower than the institutions of the third quartile. Inputs are distributed between the four quartiles without any specific rule, the enrolled student input his higher in the 2nd and 4th quartile, but quite close to the 1st and 3rd; teaching and staff inputs are substantially costant across quartiles, while PhD are increasing with the efficiency level (from the first to the fourth quartile). Finally student services measure has a u-shaped distribution across quartiles, with higher levels in first and fourth ones woth respect to second and third.

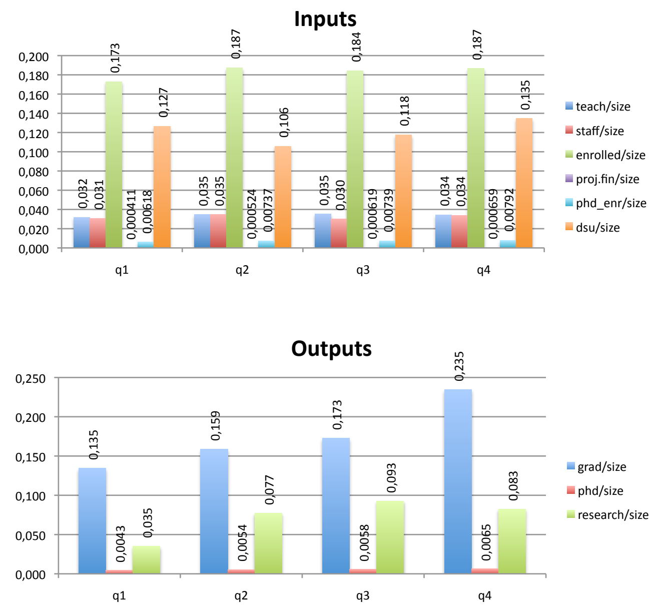


Figure 8 a) and b) - average Inputs and Outputs (size corrected) measures by efficiency quartile

As discussed in previous paragraphs we expect that the heterogeneity of italian universities means differences in efficiency levels by size and tipe of institution. A comparison of the values in Table 5 for size (small and medium/large universities), state versus private institutions and polytechnic versus the others shows that size and technical propension matter, while the type of institutional form does not. A test of the null hypothesis that means are the same for all groups are rejected in two of the three cases. Difference in efficiency means for independent groups of institutions indicate that there is a statistically significant difference between mean efficiency of medium/large universities (0.947) vs small universities (0.924) and for state against private academies (0.9355 vs 0.9350). polytechnic difference is not statistically significant.

	Freq	*avg_ Eff*	*st.dev*	*Test of Variances*	*Pr>F*	*Test of Means*	*Pr>/t/*
Medium/Large Universities	28	0.947	0.0439	Folded-F	<.0001	Satterthwaite	0.0013
Small-Universities	31	0.924	0.0640				
State Universities	53	0.9355	0.0516	Folded-F	<.0001	Satterthwaite	0.9805
Private Universities	6	0.9350	0.0905				
Polytechnic Universities	4	0.9335	0.0567	Folded-F	0.1356	Pooled	0.0267
Other Universities	55	0.9705	0.0038				

Table 5 - Average Efficiency by size, nature and type of Institution

Table six presents efficiency scores of the random paramenter stochastic frontier distance function model for each insitution. It's evident the flexible nature of the scores that can vary across institutions and over time allowing for both time and institutional heterogeneity.

University	2005	2006	2007	2008
Ancona - Università Politecnica delle Marche	0,958787	0,972596	0,925032	0,959907
Bari - Politecnico	0,956432	0,958692	0,894511	0,895688
Bari - Università degli studi	0,961725	0,970529	0,974043	0,955125
Benevento - Università degli studi del Sannio	0,794272	0,793083	0,814051	0,832286
Bergamo - Università degli Studi	0,883157	0,845281	0,894049	0,871059
Bologna - Università degli studi	0,99332	0,987701	0,985782	0,982551
Bolzano - Libera Università	0,803565	0,681247	0,72371	0,906872
Brescia - Università degli studi	0,929411	0,926037	0,919884	0,940435
Cagliari - Università degli studi	0,854377	0,855234	0,855623	0,869301
Camerino - Università degli studi	0,916948	0,884243	0,910973	0,870508
Campobasso - Università degli studi del Molise	0,905934	0,89107	0,852878	0,857979
Cassino - Università degli studi	0,812182	0,930916	0,906008	0,940529
Castellanza - Università "Carlo Cattaneo"	0,995804	0,995217	0,995493	0,994452
Catania - Università degli studi	0,839546	0,873813	0,86728	0,925197
Catanzaro - Università degli studi "Magna Grecia"	0,96295	0,957236	0,960109	0,959852
Chieti - Università degli studi Gabriele D'Annunzio	0,994728	0,994719	0,989278	0,955504
Ferrara - Università degli studi	0,969759	0,982999	0,98037	0,942549
Firenze - Università degli studi	0,968125	0,979207	0,981666	0,985458
Foggia - Università degli studi	0,800402	0,952894	0,832248	0,925142
Genova - Università degli studi	0,869258	0,862985	0,862449	0,869053
Lecce - Università degli studi	0,904694	0,926271	0,909806	0,929978
Macerata - Università degli studi	0,919323	0,905543	0,937621	0,947467
Messina - Università degli studi	0,855651	0,847025	0,871818	0,871321
Milano - Politecnico	0,9979	0,998104	0,998742	0,998173
Milano - Università Cattolica del "Sacro Cuore"	0,970736	0,937685	0,986333	0,988117
Milano - Università commerciale "Luigi Bocconi"	0,99919	0,998343	0,999497	0,999543
Milano - Università degli studi	0,918063	0,921244	0,944854	0,962087
Milano-Bicocca - Università degli studi	0,962861	0,969096	0,976442	0,969202
Modena e Reggio Emilia - Università degli studi	0,956954	0,923925	0,942449	0,949537
Napoli - Istituto Universitario "Suor Orsola Benincasa"	0,882034	0,829213	0,900973	0,93043
Napoli - Seconda Università degli studi	0,938571	0,942506	0,941352	0,,940916
Napoli - Università degli studi "Federico II"	0,905185	0,911076	0,930739	0,944308
Padova - Università degli studi	0,996342	0,994814	0,995911	0,994651

University	2005	2006	2007	2008
Palermo - Università degli studi	0,959997	0,962333	0,958518	0,917041
Parma - Università degli studi	0,977236	0,975379	0,961559	0,957069
Pavia - Università degli studi	0,97892	0,975837	0,973624	0,977038
Perugia - Università degli studi	0,974724	0,972714	0,97224	0,971258
Pisa - Università degli studi	0,965328	0,948446	0,959324	0,961
Potenza - Università degli studi della Basilicata	0,845965	0,773551	0,854377	0,865419
Reggio Calabria - Università degli studi Mediterranea	0,861457	0,864545	0,852882	0,858591
Roma - III Università degli studi	0,946806	0,932316	0,975075	0,942697
Roma - Libera Università Maria SS.Assunta (LUMSA)	0,978217	0,976802	0,982848	0,983897
Roma - Università degli studi "La Sapienza"	0,97089	0,989376	0,967383	0,989983
Roma - Università degli studi di "Tor Vergata"	0,971375	0,957592	0,906597	0,947324
Salerno - Università degli studi	0,88307	0,887915	0,875922	0,87727
Sassari - Università degli studi	0,848636	0,855521	0,825656	0,840012
Siena - Università degli studi	0,986702	0,990723	0,991375	0,994792
Teramo - Università degli studi	0,940804	0,910112	0,968135	0,93359
Torino - Politecnico	0,988137	0,987872	0,985199	0,987084
Torino - Università degli studi	0,996572	0,974365	0,986945	0,989311
Trento - Università degli studi	0,971634	0,95874	0,959123	0,964327
Trieste - Università degli studi	0,992608	0,979146	0,98479	0,985929
Udine - Università degli studi	0,932289	0,958161	0,961042	0,947283
Urbino - Università degli studi "Carlo Bo"	0,992071	0,997232	0,997495	0,981005
Varese - Università dell' Insubria	0,905665	0,947791	0,942252	0,897249
Venezia - Università IUAV	0,990867	0,992477	0,994628	0,990348
Venezia - Università degli studi "Cà Foscari"	0,938342	0,922843	0,9382	0,92993
Verona - Università degli studi	0,959989	0,959158	0,9449	0,926741
Viterbo - Università della Tuscia	0,964058	0,980087	0,980925	0,98422

Table 6 - Random Parameters SFA Efficiency scores and university rankings by year

University	2005	2006	2007	2008
Ateneo	ranking '05	ranking '06	ranking '07	ranking '08
Milano - Università commerciale "Luigi Bocconi"	1	1	1	1
Milano - Politecnico	2	2	2	2
Torino - Università degli studi	3	19	9	8
Padova - Università degli studi	4	5	4	4
Castellanza - Università "Carlo Cattaneo"	5	4	5	5
Chieti - Università degli studi Gabriele D'Annunzio	6	6	8	26
Bologna - Università degli studi	7	11	11	15
Trieste - Università degli studi	8	15	13	11
Urbino - Università degli studi "Carlo Bo"	9	3	3	16
Venezia - Università IUAV	10	7	6	6
Torino - Politecnico	11	10	12	10
Siena - Università degli studi	12	8	7	3
Pavia - Università degli studi	13	17	21	17
Roma - Libera Università Maria SS.Assunta (LUMSA)	14	16	14	14
Parma - Università degli studi	15	18	25	25
Perugia - Università degli studi	16	20	22	18
Trento - Università degli studi	17	26	29	20
Roma - Università degli studi di "Tor Vergata"	18	29	43	30
Roma - Università degli studi "La Sapienza"	19	9	24	7
Milano - Università Cattolica del "Sacro Cuore"	20	35	10	9
Ferrara - Università degli studi	21	12	17	34
Firenze - Università degli studi	22	14	15	12
Pisa - Università degli studi	23	32	28	22
Viterbo - Università della Tuscia	24	13	16	13
Catanzaro - Università degli studi "Magna Grecia"	25	30	27	24
Milano-Bicocca - Università degli studi	26	23	18	19
Bari - Università degli studi	27	22	20	27
Palermo - Università degli studi	28	24	30	45
Verona - Università degli studi	29	25	31	42
Ancona - Università Politecnica delle Marche	30	21	39	23
Modena e Reggio Emilia - Università degli studi	31	40	33	28
Bari - Politecnico	32	27	46	48
Roma - III Università degli studi	33	36	19	33

Table 7 - University rankings by year

University	2005	2006	2007	2008
Teramo - Università degli studi	34	44	23	38
Napoli - Seconda Università degli studi	35	34	35	35
Venezia - Università degli studi "Cà Foscari"	36	41	36	41
Udine - Università degli studi	37	28	26	31
Brescia - Università degli studi	38	39	40	37
Macerata - Università degli studi	39	45	37	29
Milano - Università degli studi	40	42	32	21
Camerino - Università degli studi	41	48	41	52
Campobasso - Università degli studi del Molise	42	46	55	57
Varese - Università dell' Insubria	43	33	34	47
Napoli - Università degli studi "Federico II"	44	43	38	32
Lecce - Università degli studi	45	38	42	40
Bergamo - Università degli Studi	46	55	47	51
Salerno - Università degli studi	47	47	48	49
Napoli - Istituto Universitario "Suor Orsola Benincasa"	48	56	45	39
Genova - Università degli studi	49	51	51	54
Reggio Calabria - Università degli studi Mediterranea	50	50	54	56
Messina - Università degli studi	51	54	49	50
Cagliari - Università degli studi	52	53	52	53
Sassari - Università degli studi	53	52	57	58
Potenza - Università degli studi della Basilicata	54	58	53	55
Catania - Università degli studi	55	49	50	43
Cassino - Università degli studi	56	37	44	36
Bolzano - Libera Università	57	59	59	46
Foggia - Università degli studi	58	31	56	44
Benevento - Università degli studi del Sannio	59	57	58	59

Table 8 shows the results of Spearman ranking correlations in the selected years. It's statistically significant and positive for each pair of years and around 90%. So the results are persistent over time.

Spearman Correlation	ranking 05	ranking 06	ranking 07	ranking 08
ranking 05	1.0000	0.90169 (<.0001)	0.91677 (<.0001)	0.86774 (<.0001)
ranking 06	0.90169 (<.0001)	1.0000	0.87756 (<.0001)	0.83881 (<.0001)
ranking 07	0.91677 (<.0001)	0.87756 (<.0001)	1.0000	0.89369 (<.0001)
ranking 08	0.86774 (<.0001)	0.83881 (<.0001)	0.89369 (<.0001)	1.0000

Table 8 - Spearman Correlation between rankings over time

2.6.2 Grade Inflation

In Table 9 efficiency rankings by year of the quantity-quality teaching input/output model are shown. What is clear is that the rankings are consistent from year to year as in the previous case with considerably high correlation coefficients between the considered years (Table 10). What is surprising is that the effect on rankings is huge and most of the universities in the last positions in the previous specification shift now from the bottom to the top of the ranking. This effect is measured by the correlation between these new rankings and the same rankings in the previous model confirming that there is a negative, statistically significant, correlation between them.

Ateneo	ranking2005	ranking2006	ranking2007	ranking2008
Napoli - Seconda Università degli studi	1	2	1	1
Roma - Università degli studi "La Sapienza"	2	1	3	31
Reggio Calabria - Università degli studi Mediterranea	3	6	7	8
Potenza - Università degli studi della Basilicata	4	4	52	55
Cagliari - Università degli studi	5	5	8	6
Messina - Università degli studi	6	3	4	4
Bergamo - Università degli studi	7	7	25	19
Cassino - Università degli studi	8	20	11	7
Udine - Università degli studi	9	19	16	13
Bari - Politecnico	10	12	6	5
Milano - Università degli studi	11	11	28	29
Campobasso - Università degli studi del Molise	12	10	5	3
Firenze - Università degli studi	13	18	26	25
Foggia - Università degli studi	14	13	9	22
Catania - Università degli studi	15	27	27	24
Pisa - Università degli studi	16	15	24	26
Salerno - Università degli studi	17	29	20	18
Trento - Università degli studi	18	8	17	16
Parma - Università degli studi	19	16	12	9
Pavia - Università degli studi	20	25	29	21
Modena e Reggio Emilia - Università degli studi	21	17	18	12
Torino - Politecnico	22	23	21	20
Benevento - Università degli studi del Sannio	23	45	34	42
Perugia - Università degli studi	24	21	23	17
Bari - Università degli studi	25	26	32	28
Teramo - Università degli studi	26	9	13	10
Lecce - Università degli studi	27	31	22	44
Brescia - Università degli studi	28	22	30	23
Bologna - Università degli studi	29	28	31	27
Ancona - Università Politecnica delle Marche	30	30	14	33
Torino - Università degli studi	31	24	36	30
Catanzaro - Università degli studi "Magna Grecia"	32	36	43	38
Camerino - Università degli studi	33	40	15	15

University	2005	2006	2007	2008
Varese - Università dell' Insubria	34	39	40	37
Milano - Università Cattolica del "Sacro Cuore"	35	14	37	32
Roma - Università degli studi di "Tor Vergata"	36	37	19	35
Viterbo - Università della Tuscia	37	34	46	51
Milano-Bicocca - Università degli studi	38	43	42	39
Ferrara - Università degli studi	39	41	41	40
Milano - Politecnico	40	42	44	43
Siena - Università degli studi	41	35	45	48
Padova - Università degli studi	42	38	38	36
Napoli - Università degli studi "Federico II"	43	33	10	11
Roma - III Università degli studi	44	46	50	46
Venezia - Università degli studi "Cà Foscari"	45	44	47	47
Trieste - Università degli studi	46	32	35	41
Castellanza - Università "Carlo Cattaneo"	47	48	48	50
Napoli - Istituto Universitario "Suor Orsola Benincasa"	48	49	49	49
Urbino - Università degli studi "Carlo Bo"	49	47	57	45
Sassari - Università degli studi	50	50	2	2
Bolzano - Libera Università	51	51	53	53
Venezia - Università IUAV	52	53	51	54
Roma - Libera Università Maria SS.Assunta (LUMSA)	53	54	58	57
Palermo - Università degli studi	54	57	39	14
Milano - Università commerciale "Luigi Bocconi"	55	52	55	56
Chieti - Università degli studi Gabriele D'Annunzio	56	55	56	58
Macerata - Università degli studi	57	56	54	52
Genova - Università degli studi	58	58	59	59
Università degli studi di Verona	59	59	33	34

Table 9 - University rankings using quantity-quality measures of teaching inputs and outputs

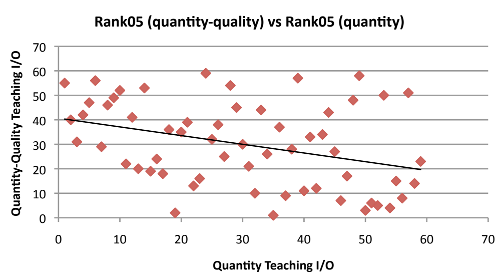
Spearman Correlation	ranking 05	ranking 06	ranking 07	ranking 08
ranking 05	1.0000	0.92466 (<.0001)	0.71479 (<.0001)	0.64980 (<.0001)
ranking 06	0.92466 (<.0001)	1.0000	0.70900(<.0001)	0.65149 (<.0001)
ranking 07	0.71479 (<.0001)	0.70900(<.0001)	1.0000	0.89936 (<.0001)
ranking 08	0.64980 (<.0001)	0.65149 (<.0001)	0.89936 (<.0001)	1.0000

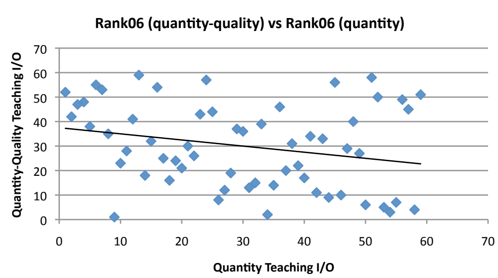
Table 10 - Spearman Correlation between q-q* rankings over time

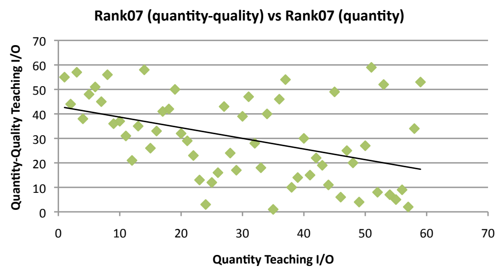
The table below (Table 11) reportes the Spearman correlation values, null hypotesis of zero correlation is rejected (p-value in brackets) and so there is a slighty negative correlation of 0.35 between the two rankings referring to the same academic year. Figures 9 plot the two alternative rankings showing that for each of the considered years the relationship is slighty, and negative.

Spearman Correlation	ranking q-q*	ranking q*
ranking q-q*	1.00	-0.35517 (0.0058)
ranking q*	-0.35517 (0.0058)	1.00

Table 11 - Spearman Correlation coefficients between q-q* and q* rankings







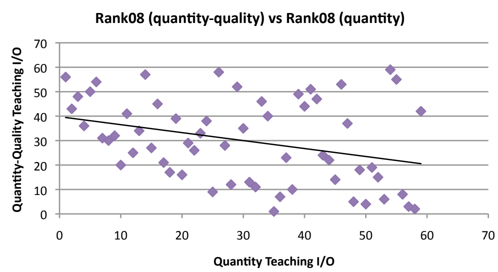
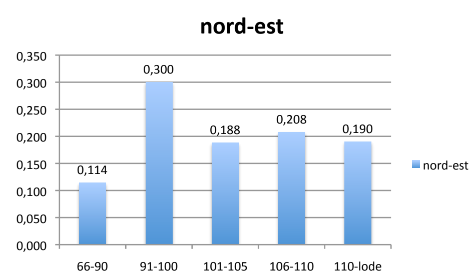


Figure 9 a) b) c) d) - quantity-quality i/o teaching ranking vs quantity i/o teaching ranking

It's self-evident for everyone, both pratictioners and starters, that new rankings are unrealistic in the Italian case; Universities of Potenza, Messina, Reggio Calabria, Campobasso, Foggia and the others in the top of table 9 are always in bottom positions in institutional evaluation exercises (Civr and Cnvsu sources). In order to identify if there is a real misleading effect in this table we have fitted the same stochastic frontier random parameters model excluding alternatively each input and output out from the production function. Results shown that the ranking does not suffer of relevant changes except for the exlusion of quantity-quality teaching input (totqual) and output (gradqual). Substituting these with their absolute values (corrected by size and mean) the new ranking table obtained is Table 7.

Is there something wrong in quality data or not? Usually if grades are inflated the grade distribution is shifted towards better grades and becomes compressed at the upper tail. Looking at the histograms in figure 10 (a-...-d) it is clear that the distribution of graduates by marks (divided in 4 classes) is different by macro-region. Universities in the north-est and north-west parts of the country show quasi-normal distributions while in middle and south regions the distrbition is shifted to the right with more than 45% of graduates obtaining a mark greater than 106/110. Tha last graph (figure 11) shows that distribution of marks is lower for the first-third classes and higher in last two intervals for middle and south universities with respect to the northen ones (both in east and west northen parts). Our graphs cannot give more detailed informations beacause of the few number of mark intervals available. What is clear is that in these regions students get better grades. Two possible explanations are given to this phenomenon: concentration of better students is particoularly nested in central and south regions or achievement requirements for passing exams and obtain a degree are decreasing in these universities. The general opinion is that the former hypothesis is unrealistic and so the latter is more realistic. Indeed it's well known in Italian Education literature at all educational levels (starting from Invalsi test in primary school up to Ocse-Pisa results in the high schools) that our system suffer from a severe regional-bias on grades. Average grades are higher in south and middle regions with respect to the northen ones. Support to this is given by other sources of institutional information about achievement evaluation in Italy for lower educational levels. It is a well-known problem for example in Invalsi standardized tests of shool achievement where the own evaluating institution is studying a strategy to overcome this problem and correct the scores distribution by this problems (Figures 17 in Appendix). In addition to this while the Ocse-Pisa results of standardized test on Italian high-school students (in 2006) shows that students from southern regions in Italy are in the bottom part of the Ocse leauge table (and students from Italian central and northen regions get higher positions) Italian institutional data about high-school marks (source Miur) in the same year assess the high percentage of top marks in these regions.



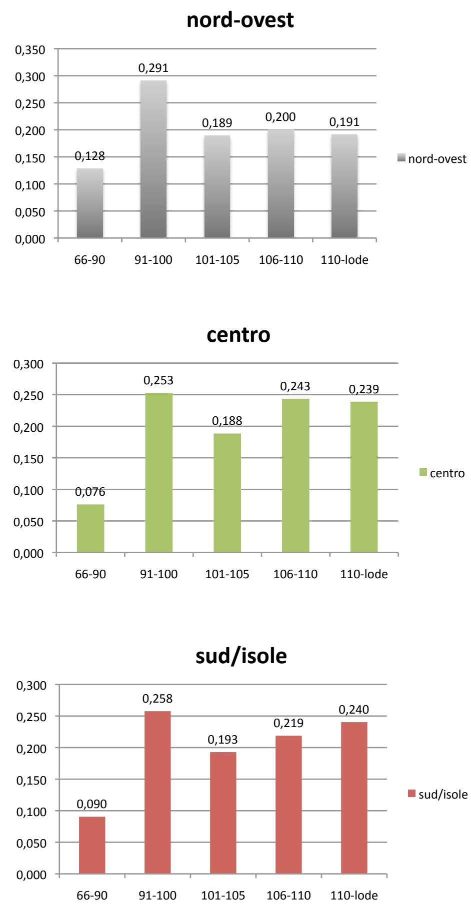


Figure 10 a) b) c) d) - 2008 graduates distribution by grade intervals in the forth macro-regions.

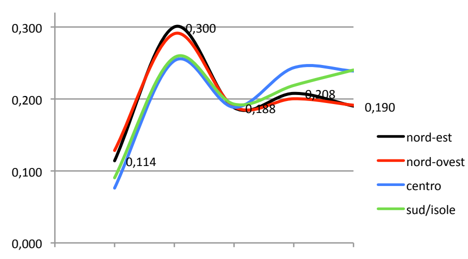


Figure 11 - 2008 graduates distribution by grade intervals

There are some possible explanantions to grade inflation in Italian Universities. My belief is that we have to consider two possible sources of grade-bias in the University system and one probable future negative effect. The first source is due to institution (teaching and professor staff) unintentional cheating, which means that universities in the south or middle regions are not able to determine in an approriate way achievement and abilities of their students. Grades does not provide a realistic metric as proxy of student's ability. The second source could be a realistic explanation of the first. Due to labor-market system features in the different regions of the country and higher young unemployment rate, longer university-to-work transitions and lower average incomes are recorded in south and middle regions; this issue probably makes grade inflation a possible, in most of the cases also unintentional, answer to students' working, income and life expectations.

Finally the probable negative future effect is structure-related at system level. As a matter of fact with the introduction of indicator-based funding system (Law n° 1, 2009) it's probable that grade-inflation could be a severe problem for Italian Universities. These financing system aimed at increasing the competition between universities by making a share (7% in 2009, 10% in 2010, up to 30% in 2013) of their pubic funds dependent on their relative performances concerning different output measures, such as the share of students obtaining a degree, their time to degree and the amount of third party funds. In contrast to traditional funding systems in which the budget of a given year was basically related on past budget values, the new funding system takes the actual performance of a university into account for an increasing share of the total amount of fundings. It is possible, and in some sense probable, that this system pushes universities to inflate grades in order to get more funds. Existing evidence shows that a funding system that concentrates on a few output indicators may lead to wrong incentives. Using for example the number of graduates as an indicator to determine the yearly amount of public funds may reduce quality standards to increase the amount of graduates rather than increasing teaching quality. In such a case twice is the negative effect, the reform may increase grade-inflation and on the other side the goal of increasing teaching quality due to the reform is not reached (Bauer, T.K. and Graves, B.S., 2011).

2.7 Conclusions

This paper has provided a random parameters approach which might be used to assess efficiency in Italian Higher Education Institutions. Our SFA model is applied in a multiple-input multiple-output production context. The application of this model is illustrated using a data set of Italian universities in period 2005-2008.

Our study should be considered as a novelty in Italian economic literature thanks to the methodological approach and the data avilability. We propose a translog distance function approach to measure techincal efficiency of Italian universities under a recent evolution of the stochastic frontier analysis. The recent version of Random Parameters stochastic frontier models is fitted in order to allow institutional heterogeneity in efficiency estimations over time. Direct measures of research output both referring to their quality and quantity are provided by ISI-Thompson datasets and Journal Citations Report.

An examination of the input and output composition of the highest and lowest sets of HEIs with respect to their efficiency suggests that the level of inputs does not vary so much across institutions quartiles but the most efficienct universities tend to be different in outputs composition. The most efficients tend to be largest producers (on average) of graduates and of PhD students, but not absolutely the largest producers by quantity and quality of their reserach products (3rd quartiles institutions are better).

A strong caveat must accompany our results, even if the model incorporate a subject-mix control variable in order to disentangle the effect of subject-mix composition of universities on the reserach ISI output, we are consciouss of the fact that in our future work, waiting for direct trustworthy indexes of italian research output (quantity of research products) and related measures of their quality, we need to consider a more disaggregated level of inputs and outputs to study *ceteris paribus* the technical efficiency in a departmental or faculty perspective.

A rising and not yet resolved problem is the accountability of grade-inflation effect in Italian Universities. Both quality measures of inputs and outputs of teaching production are grade-inflated towards central and southern universities, with strong differences in grade distributions across macro regions. It's self evident that institutional scores for high schools and university courses cannot be considered as ralistic measures of the student performances. Incorporating measures of quality particularly in the context of teaching inputs and outputs has a misleading effect on efficiency estimations and efficiency rankings. It's

Therefore a study of substituability between inputs and outputs (using for example Morishima elasticities; Jhones, 2010) was outside our goals in this paper, but could be an interesting perspective of future studyies in order to analyze the degree of substitution between teaching, research and managerial inputs in order to obtain greater efficiency levels.

Another future extension of this work is the inclusion in the production function for HEIs in addition to reserach, teaching and managerial dimensions, of the 'third mission' of the University System, the so called 'social mission'. This includes such services provided by Universities as the storage of knowledge, the provision of advice to business and firms and its own public interests dimension. This measure of output is really rare in literature due to the availability of quantitative realistic information over them; results of our work may be biased as a consequence of the omission of this social dimension (Johnes, 2010). Further work in this area should also include comparative studies across countries (matching our work with the Spanish and UK efficiency literature for example) in order to abtain feasible comparative results across countries.

References

1. Athanassopoulos,A.D., and Shale, E., (1997), 'Assessing the comparative efficiency of higher education institutions in the UK by means of data envelopment analysis', Education Economics 5 (2), pp. 117-134;

2. Agasisti, T. and Johnes, G., (2010), 'Heterogeneity and the evaluation of efficiency: the case of Italian universities', Applied Economics, 42: 11, pp. 1365-1375;

3. Aghion, P., Dewatripont, M., Hoxby, C.M., Mas-Colell, A. and Sapir, A., (2009), 'The Governance and Performance of Research Universities: Evidence from Europe and the U.S.', NBER Working Paper n. 14851;

4. Aigner, D.J., Lovell, C.A.K., Schmidt, P., (1977), 'Formulation and estimation of stochastic frontier production function models'. Journal of Econometrics 6, pp. 21-37;

5. Atkinson, S.E., and Dorfman,J. H. , (2005), 'Bayesian Measurement of Productivity and Efficiency in the Presence of Undesirable Outputs: Crediting Electric Utilities for Reducing Air Pollution', Journal of Econometrics 126, (2): 445-468;

6. Atkinson, S.E., Fare, R. and Primont, D. , (2003), 'Stochastic estimation of firm inefficiency using distance functions'. Southern Economic Journal 69(3): 596-611;

7. Atkinson, S.E., Cornwell, C. and HonerKamp, O. , (2003), 'Measuring and decomposing productivity change: stochastic distance function estimation vs DEA', Journal of Business and Economic Statistics 21(2): 284-294;

8. Atkinson, S.E., Primont, D. , (2002), 'Stochastic estimation of firm technology, inefficiency, and productivity growth using shadow cost and distance functions. 'Journal of Econometrics', 108: 203-225;

9. Avkiran, N. K., (2001), 'Investigating technical and scale efficiencies of Australian universities through data envelopment analysis', Socio-Economic Planning Sciences, 35: 57-80;

10. Balcombe, K., Doucouliagos, H., and Fraser, I., (2007), 'Input usage, output mix and industry deregulation: an analysis of the Australian dairy

manufacturing industry', The Australian Journal of Agricultural and Resource Economics, 51(2): 137-156;

11. Banos-Pino, J., Fernandez-Blanco, V. and Rodrìguez-Alvarez, A. (2002), 'The allocative efficiency measure by means of a distance function: the case of Spanish railways', European Journal of Operational Research, 137(1): 191-205;

12. Battese, G. E. and Coelli, T. J. (1988), 'Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data', Journal of Econometrics, 38, pp. 387-399;

13. Battese, G. E. and Coelli, T. J. (1995), ' A model for technical inefficiency effects in stochastic frontier production function for panel data'. Empirical Economics 20, pp. 325-332;

14. Bauer, T.K., Graves, B.S. (2011), 'Performance-related Funding of Universities: Does More Competition Lead to Grade Inflation?', IZA Discussion Paper No. 6073;

15. Baumol, W. J., Panzar, J. C. and Willig, R. D. (1982), 'Contestable Markets and the Theory Of Industry Structure', Harcourt Brace Jovanovich, San Diego;

16. Breno, E. et al., (2002), "Scientific Research in Italian Univesities: an initial analysis of the citations in the ISI data bank", CRUI;

17. Cave M., Weale M., (1992), 'The Assessment: Higher Education: The State of Play', Oxford Review of Economic Policy, 8, pp. 1-18;

18. Coelli, T.J., Rao, D.S., and Battese, G.E., (1999), 'An Introduction to Efficiency and Productivity Analysis', Kluwer Academic Publishers: North Holland;

19. Coelli, T.J., (2000), 'On the econometric estimation of the distance function representation of a production technology', Working Paper 2000/042;

20. Coelli, T. and Perelman, S. (1999), 'A comparison of parametric and non-parametric distance functions: with application to European railways'. European Journal of Operational Research, 117, pp. 326-339;

21. Coelli, T. and Perelman, S. (2000), 'Technical efficiency of European railways: a distance function approach', Applied Economics, 32: 1967-1976;

22. Cohn, E., Cooper, S., T. (2004), 'Multi-product cost functions for universities: economies of scale and scope', The International Handbook on the Economics of Education, G. Johnes and J. Johnes (ed). Cheltenham, Edward Elgar;

23. Cohn, E., Rhine, S.L.W., Sanros, M.C., (1989), 'Institutions of higher education as multi-product firms: economies of scale and scope', Review of Economics and Statistics, 71(2): 284-290;

24. Cuesta, R. A., and Orea, L., (2002), 'Mergers and technical efficiency in Spanish savings banks: a stochastic distance function approach', Journal of Banking and Finance, 26(12): 2231-2247;

25. Cuesta, R. A., and Zofío, J., (2005), 'Hyperbolic efficiency and parametric distance functions:with application to Spanish savings banks', Journal of Productivity Analysis, 24: 31-48;

26. Econometric Software, inc. (2002). LIMDEP, Version 8.0. ESI, New York;

27. Fecher, F., D. Kessler, S. Perelman, and P. Pestieau, (1993), "Productive Performance of the French Insurance Industry," Journal of Productivity Analysis 4: 77-93;

28. Ferrier, G., and Lovell, C.A.K., (1990), "Measuring Cost Efficiency in Banking: Econometric and Linear Programming Evidence", Journal of Econometrics 46, 229–245;

29. Glass, J.C., McKillop, D.G., and O'Rourke, G., (1998), 'A cost indirect evaluation of productivity change in Uk universities', Journal of Productivity Analysis 10(October): 153-175;

30. Green, A., Mayes, D., (1991), 'Technical inefficiency in manufacturing industries'. Economic Journal 101, pp. 523–538;

31. Greene, W. (2005), 'Reconsidering heterogeneity in panel data estimators of the stochastic frontier model', Journal of Econometrics, 126, pp. 269–303;

32. Hashimoto, K., Cohn, E., (1997), 'Economies of scale and scope in Japanese private universities', Education Economics 5(2):107-115;

33. Izadi, H., Johnes, G., Oscrochi, R., and Crouchley, R., (2002), 'Stochastic frontier estimation of a CES cost function: the case of higher education in Britain', Economics of Education Review, 21, pp. 63-72;

34. Johnes, G., (1992), 'Performance Indicators in Higher Education: A Survey of Recent Work', Oxford Review of Economic Policy, 8, pp. 19-34;

35. Johnes, G., (1996), 'Multi-product cost functions and the funding of tuition in Uk universities', Applied Economics Letters, 3(9): 557-561;

36. Johnes, G., (1997), 'Costs and industrial structure in contemporary British higher education', Economic Journal, 107, pp. 727–37;

37. Johnes, G., and Johnes, J. (1993), 'Measuring the research performance of UK economics departments', Oxford Economic Papers, 45, pp. 332–47;

38. Johnes, G., and Johnes, J. (2006), 'Higher education institutions costs and efficiency: taking the decomposition a further step', Economics of Education Review, 28(1):107-113;

39. Johnes, J., (2006), 'Data Envelopment Analysis and its application to the measurement of efficiency in higher education', Economics of Education Review,25, pp. 273-288;

40. Johnes, J., and Taylor, J., (1997), 'Degree Quality: An Investigation into Differences between UK Universities', Higher Education, 16, pp. 581-602;

41. Johnes, J., (2010), 'Efficiency and Input-Output-Substitutionability in English Higher Education 1996/97 to 2007/08: A Parametric Distance Function Approach', Lancaster University Management School Workin paper 2010/025;

42. Jondrow, J., Lovell, C. A., Knox, M., Ivan, S. and Schmidt, P. (1982), 'On the estimation of technical inefficiency in the stochastic frontier production function model', Journal of Econometrics, 19, pp. 233–8;

43. Karagiannis, G., Midmore, P., and Tzouvelekas, V., (2004), 'Parametric decomposition of output growth using a stochastic input distance function', American Journal od Agriculture Economics, 86(4): 1044-1057;

44. Koshal, R.K., and Koshal, M., (1999), 'Economics of scale and scope in higher education: a case of comprehensive universities', Economics of Education Review, 18, pp. 270- 7;

45. Kumbhakar, S., (1990), 'Production frontiers and panel data and time varying technical efficiency'. Journal of Econometrics 46, pp. 201-211;

46. Kumbhakar, S., Gosh, S., and McGulkin, J.T., (1991), 'A Generalized Production Frontier Approach for Estimating Determinants of Inefficiency in U.S. Dairy Farms', Journal of Business & Economic Statistics, American Statistical Association, vol. 9(3), pp. 279-86;

47. Mandelson, P., (2009), 'Higher Education Funding 2010-11', Letter to HEFCE (Retrieved from http://www.bis.gov.uk/wp-content/uploads/publications/Mandelson-Letter-to-HEFCE- Dec09.pdf);

48. Marra, A., (2009), 'Dall'uniformità alla differenziazione ovvero ripensare l'università italiana'. Intervento presentato a: Il governo dei giudici? Dialogo tra le corti europee e nazionali, Barcellona, Spagna;

49. Meeusen, W., and van den Broeck J., (1977), 'Efficiency Estimation from Cobb-Douglas Production Functions With Composed Error', International Economic Review, 18, pp. 435-44;

50. O'Donnell, C.J., and Coelli, T.J., (2005), 'A Bayesian approach to imposing curvature on distance functions', Journal of Econometrics, Elsevier, vol. 126(2), pp. 493-523;

51. Paul, C. J. M., Johnston,W. E. , and Frengley, G.A.G., (2000), 'Efficiency in New Zealand sheep and beef farming: the impacts of regulatory reform', The Review of Economics and Statistics, 82(2): 325-337;

52. Paul, C. J. M., Johnston,W. E. , and Frengley, G.A.G., (2002), 'Post-reform substitution and cost efficiency in the New Zealand agricultural sector', International Journal of Business and Economics, 1(2): 135-146;

53. Paul, C. J. M., and Nehring, R., (2005), 'Product diversification, production systems, and economic performance in US agricultural production', Journal of Econometrics, 126: 525-548;

54. Perelman, S., and Santìn, D., (2011), 'Measuring educational efficiency at student level with parametric stochastic distance functions: an application to Spanish PISA results', Education Economics, Taylor and Francis Journals, vol. 19(1), pages 29-49;

55. Pitt, M., Lee, L., (1981), 'The Measurement and Sources of Technical Inefficiency in Indonesian Weaving Industry', Journal of Development Economics, 9, 43-64;

56. Reifschneider, D., and Stevenson, R., (1991), 'Systematic departures from the frontier: A framework for the analysis of firm inefficiency'. International Economic Review, 32 (3): 715-723;

57. Rodríguez-Álvarez, A., Fernández-Blanco, V., and Lovell, C. A., (2004), 'Allocative inefficiency and its cost: the case of Spanish public hospitals', International Journal of Production Economics, 92: 99-111;

58. Saal, D. S., and Parker, D., (2006), 'Assessing the performance of water operations in the English and Welsh water industry: a panel input distance function approach', Performance Measurement and Regulation of Network Utilities, T. Coelli and D. Lawrence (ed), Cheltenham, Edward Elgar: pp297-328;

59. SAS Institute Inc., SAS 9.1.3, Cary, NC: SAS Institute Inc., 2002-2004;

60. Sarrico, C.S., Dyson, R.G., (2004), 'Restricting virtual weights in data envelopment analysis', European Journal of Operational Research 159 (1), 17–34;

61. Sarrico, C.S., Teixeira, P.N., Rosa, M.J., Cardoso, M.F., (2009), 'Subject mix and productivity in Portuguese universities', European Journal of Operational Research, 197, pp. 287–295;

62. Shepard, R. W., (1970), 'Theory of Cost and Production Functions', Princeton University Press, Princeton, NY;

63. Stevens, P. A., (2005), 'The determinants of economic efficiency in English and Welsh universities', Education Economics, 13(4), pp. 355–74;

64. Stevens, P.A., Vecchi, M., (2002), 'The effect of environmental variables on public service efficiency', Presentation to the British Academy, 19th November 2002;

65. Thanassoulis, E., Boussfofiane, A., Dyson, R G., (1995), 'Exploring output quality targets in the provision of perinatal-care in England using data envelopment analysis', European Journal of Operational Research, Vol. 80, pp. 588-607;

66. Tonini, A., (2004), 'A distance function approach for the Polish agricultural sector using generalised maximum entropy with linear inequality restrictions', Working paper;

67. Tsionas, E. G., (2002), 'Stochastic frontier models with random coefficients', Journal of Applied Econometrics, 17, pp. 127–47;

68. Uri, N. D., (2003), 'The adoption of incentive regulation and its effect on technical efficiency in telecommunications in the United States', International Journal of Production Economics, 86: 21-34;

69. Uri, N. D., (2003), 'Technical efficiency in telecommunications in the United States and the impact of incentive regulation', Applied Mathematical Modelling, 27: 53-67;

70. Vandenberghe, V., (1999), 'The need to go beyond human capital theory and production-function analysis. Educational Studies', Economics of Education, 25 (2), pp. 129-143;

71. Waldman, D.M., (1982), 'A stationary point for the stochastic frontier likelihood'. Journal of Econometrics 18, pp. 275–279;

72. Whiteman, J. L., (1999), 'The measurement of efficiency where there are multiple outputs', Centre of Policy Studies/IMPACT Centre Working Papers;

73. Worthington, A.C., (1999), 'Performance indicators and efficiency measurements in public libraries', Australian Economic Review, 32, pp.31-42;

74. Worthington, A. C., (2001), 'An empirical survey of frontier efficiency measurement techniques in education'. Education Economics, Vol. 9, No 3;

75. Worthington, A. C., and Higgs, H., (2008), 'An empirical note on factor price elasticities and substitution in Australian higher education', SSRN eLibrary;

Appendix

Some descriptive statistics, more detailed graphs and robustness checks are reported here as allowance for interested readers in better understanding econometric details. For each of the following graphs and tables is specified a brief tagline.

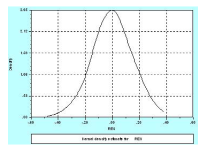


Figure 12 - Kernel density of the OLS residuals (negatively skewed)

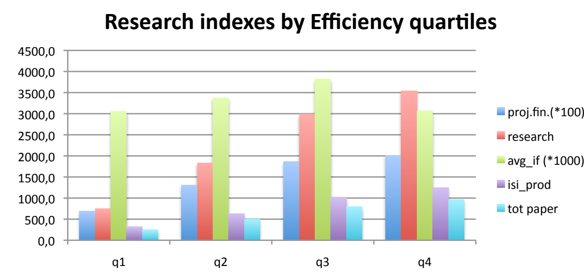


Figure 13 - Research Indexes (both input and outputs) by efficiency quartiles

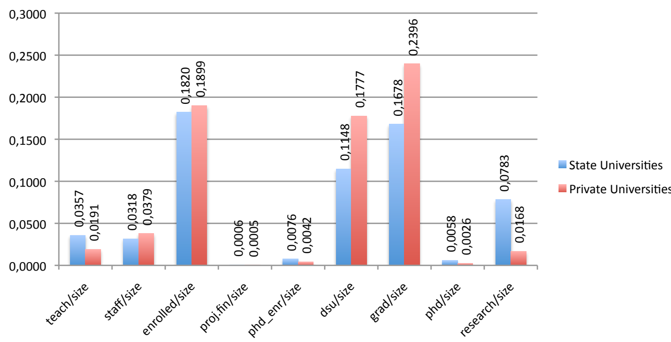


Figure 14 - Inputs and Outputs (size corrected) average measures by Type of University

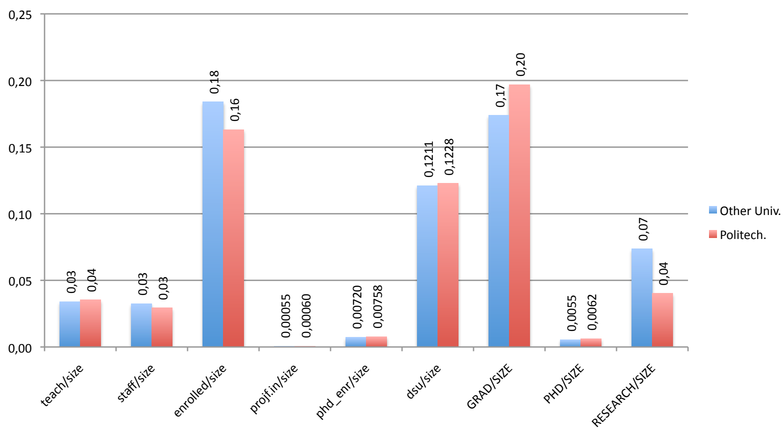
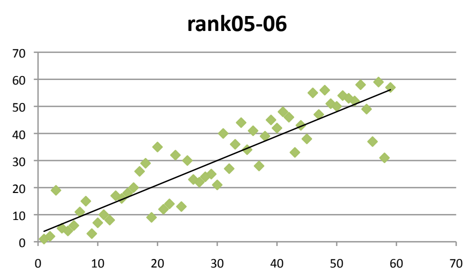
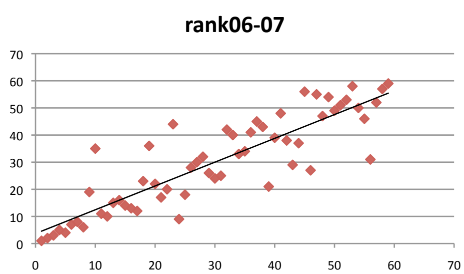


Figure 15 - Inputs and Outputs (size corrected) by polytechnic vs others





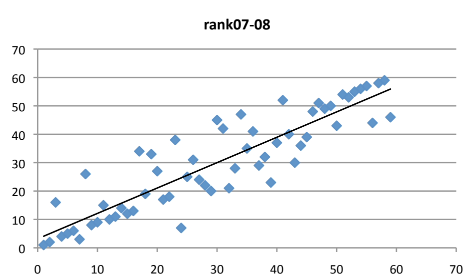


Figure 16 - Plot of RPM Rankings by pair of years

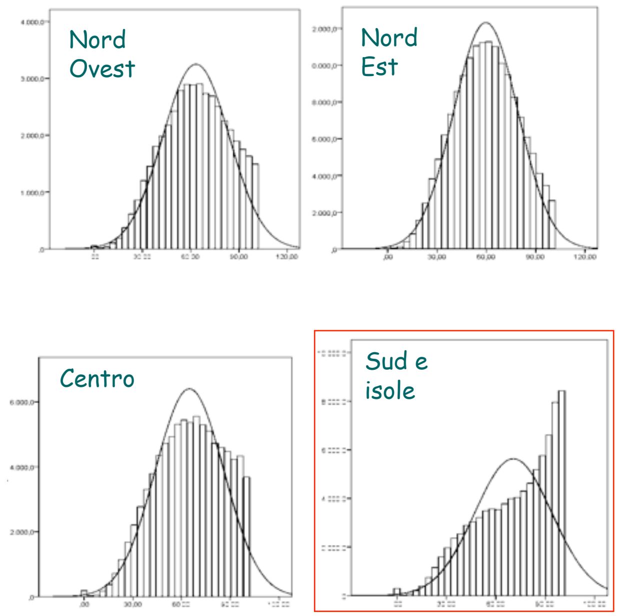


Figure 17- Math test scores distributions in Italian primary school by macro area, s.y. 2005-06, source Invalsi

3 Selecting University Professors in Italy: much ‘ado’ about nothing?

Abstract

The aim of this paper is to test the effect of decentralizing recruitment mechanisms on the average quality of researchers in the Italian academia. Quality is measured via some bibliometric indicators collected from the web version of ISI Web Of Knowledge over the last two decades (1991-2010). We test the presence of negative effects on international research quality of selected researchers in terms of quantity, impact and notoriousness of their published research products. A by product of this work is the analysis of changes in promotion criteria adopted by national and local selection committees. We found contrasting results: an overall general worsening effect of decentralization on the quality of recruited is not clearly identified. However differentiating our analysis by disciplinary area we found negative effects onto “bibliometric disciplines” with stronger effects in few disciplines (e.g. Physics, Chemistry, Biology) with respect to the left-overs. A general increase in research outcomes variability of hired researchers and a polarization effect of the selecting criteria adopted by the selection committees (where better candidates are more likely to be selected into higher quality departments with respect to national competitons) are associated with the local reform.

JEL Classification: Keywords: treatment effect, research quality, ISI Web of Knowledge

3.1 Introduction

The third and fourth chapters of this thesis are strictly related and we suggest to read them consequently. The underlying idea of both these works is to analyze if recruitment mechanisms of academic professors matters on scientists research productivity (Allesina, S., 2011; Checchi, D., 1999). Particularly in what follows we aim to verify whether the introduction of local selection procedures after the decentralization reform (Berlinguer’s reform in 1998) in Italy influences the research quality of selected associates and full professors (Levin, S. 1991 and Noser et al. 1996). As in the fourth chapter we build up a new dataset of international research outputs for the Italian professors since 1991 thanks to the web version of ISI Web of Knowledge and the individual administrative dataset of Italian professors of the Italian Ministry of University and Research (MIUR). The final, consistent, database contains 963.181 international research publications of Italian academics over the last twenty years. Some bibliometric measures of research productivity (mainly on quantity, notoriousness and impact of the published research) are calculated for each professor on the selected time period.

Even if the issue of familiar networks and it's relations with labour market outcomes, wages and school enrolment has been recently studied (e.g. Angelucci et al, 2010), lower attention has been given to educational recruitment reforms (Fox, M. F., 1983). According to our knowledge only the paper of Labartino et al. (2011) explores the effects of this decentralizing reform on recruitment procedures in terms of increasing nepotism and familism phenomena. The authors found evidences of higher probability of these misbehaviours in those territorial areas where a lower level of civic capital is present. According to our knowledge no other papers - up to now- with similar research questions are present in the Italian context.

We tested the presence (or not) of a negative effect of local selections on the quality of Italian professors (both for associates and full professors) at the time of their selection.

The paper is structured as follows: the next section introduces the issue of measuring the quality of Italian professors using bibliometric indicators and shows some descriptive statistics of the dataset we used. Then we introduce the two different recruitment procedures (local and national). The subsequent section reports both the empirical strategy and results we obtain in testing differences between pre and after reform selected scientist showing evidences of an increase of variability in international research productivity of locally recruited. We then move deeper in the analysis of the selection processes focussing on changes in the determinants of being selected with the two regimes in the last section. Conclusions and policy issues complete the paper.

3.2 Measuring the quality of professors by means of bibliometric indexes

University professors perform different tasks, among which the main one are teaching (at both undergraduate and postgraduate), research and administration. The quality of the performance in each task is difficult to measure because of unobservability of individual effort and talent, and university administrators (deans, provosts, rectors) have to rely on observable proxies, which are related to the outputs of these activities.

Quality of undergraduate teaching is difficult to measure, since the final outcome (typically in terms of future wage and/or employment prospect) is often the joint result of the collective effort of faculty members. In principle one could exploit variations in the exposure to different teachers[4]. However the students assessment is often plagued by exogenous source of variations (like gender, age, ethnicity of both the instructor and the students), and should be accounted properly. As far as postgraduate teaching, the quality of supervisors indirectly measured by supervisee, even if again confounding factors should be appropriately accounted for (in particular considering self-sorting of students into universities, as well as journal networks).

Quality in research can be measured by different indicators: publications, citations, funding, editorial boards, prizes. All these variables are potentially correlated, and are available at different degrees in different subjects and countries. In the sequel we use bibliometric measures[5] from the web version of ISI (Klavan and Boyak, 2007), since it dispenses of subjective judgments of the scholar, and it benefits from the property of cardinality (and therefore interpersonally comparable, at least within the discipline).

We are fully aware of the potential limits of a bibliometric approach (Seglen, 1997). In a sum its main drawbacks consist of:

a) it relies on the existence of large database, which are typically available for large academic communities, open to international competition;

b) as a consequence, it penalises national academic communities, which often write in their native languages and are not necessarily open to English writing and publishing;

[4] Michela Braga, Marco Paccagnella e Michele Pellizzari. 2011. Evaluating students' evaluations of professors. Banca d'Italia Tema di discussione n. 825, ottobre 2011

[5] Silvia Salini. 2012. An introduction to bibliometrics. mimeo

c) we are also aware that the diffusion of the use of bibliometric indicators in the process of research assessment induces mainstream compliance in the research community, in the attempt to publish in the top journals of each field of research[6].

Nevertheless we hold the view that the *pros* exceed the *cons* in the present case, and therefore we proceed with the use of bibliometric analysis, in the Italian academic communities where we deem it applicable.

Let us start presenting our main output variable, which consists of the number of ISI-Web of science records associated to each professor (assistant, associate or full) working in the Italian academia over the sample period (1991-2010)[7]. In table 1 we report the number of professors by appointment type and their presence in the ISI-WOS database, while in table 2 we show the yearly productivity. This gives us a clear picture that the Italian academy has experienced a rising trend in productivity recorded in WOS over the last 20 years.

This creates two order of problems in our following analysis:

a) our measure of productivity are clearly trended, and this may not be only the reflection of Italian professors doing more research, but simply the result of WOS extending its coverage over scientific journal;

b) some research areas do publish on journals that are not surveyed in ISI-WOS and for this reason they are excluded by construction;

[6] Alberto Baccini. 2010, Valutare la ricerca scientifica, il Mulino

[7] A detailed description of the creation of the dataset is in Chapter. 4, par. 4.4

Year	Assistants	Associates	Full Prof.	Tot.	% Assist*	% Assoc.*	% Full Prof.*	% Total*
1991	11750	14042	15642	41434	26.67	30.24	32.39	29.50
1992	11804	16746	14964	43514	28.92	35.71	37.50	33.86
1993	11876	17084	15739	44699	32.22	39.08	41.17	37.22
1994	13288	15915	16694	45897	35.41	41.50	44.48	40.15
1995	14011	16313	18417	48741	38.26	43.58	46.43	42.39
1996	13719	16093	19583	49395	40.80	46.07	49.13	44.83
1997	13399	15675	20105	49179	43.45	48.98	52.01	47.54
1998	13098	18108	18748	49954	44.88	52.43	55.18	50.32
1999	12905	18069	19815	50789	46.63	54.32	57.04	52.01
2000	14411	16615	19200	50226	46.66	55.76	58.59	53.09
2001	16901	17879	20255	55035	47.68	56.46	60.04	54.33
2002	18148	18504	21055	57707	48.78	57.74	60.88	55.46
2003	17997	18115	20577	56689	52.98	59.61	62.37	58.08
2004	18062	18094	21341	57497	53.99	60.89	63.60	59.18
2005	19296	18982	22186	60464	53.74	61.02	64.37	59.42
2006	19843	19084	23355	62282	54.31	62.06	65.55	60.26
2007	19640	18776	23793	62209	57.03	63.52	66.66	62.03
2008	18929	18253	25923	63105	57.80	65.20	68.03	63.01
2009	17980	17630	25911	61521	60.72	67.25	69.52	65.17
2010	15949	16967	25590	58506	62.32	69.01	70.84	66.58
2011	15231	16520	26152	57903	62.03	70.15	71.81	66.92
Tot.					49.02	54.83	58.59	53.68

Table 1 - Professors and ISI web of science publications; * percentage of professors with at least 1 ISI product

	Assistants	Associates	Full Prof.	Total
1991-2	0.437	0.607	0.818	0.606
1992-3	0.498	0.649	0.918	0.667
1993-4	0.530	0.636	0.970	0.694
1994-5	0.604	0.697	1.035	0.759
1995-6	0.674	0.758	1.140	0.831
1996-7	0.705	0.800	1.226	0.877
1997-8	0.737	0.892	1.295	0.939
1998-9	0.782	0.931	1.313	0.970
1999-0	0.778	0.940	1.380	1.004
2000-1	0.855	1.048	1.522	1.23
2001-2	0.823	1.072	1.541	1.129
2002-3	0.896	1.138	1.616	1.200
2003-4	0.941	1.197	1.667	1.250
2004-5	0.952	1.265	1.780	1.315
2005-6	1.028	1.346	1.900	1.403
2006-7	1.125	1.443	2.025	1.505
2007-8	1.1278	1.535	2.089	1.555
2008-9	1.393	1.760	2.380	1.786
2009-10	1.333	1.744	2.377	1.737
2010-11	1.274	1.648	2.228	1.632
Average	0.918	1.115	1.620	1.186

Table 2 - Average yearly productivity by level of appointment - ISI web of science publications

For these reasons, we have decided to set a minimum threshold of significance for disciplinary areas, and therefore we will consider the measure from ISI WOS as significant for scientific productivity only when the coverage exceeds a minimum threshold of 50%. Looking at table 3 we see that only eight research areas satisfy this requirement, and we will focus mainly on them.

Scientific area	% with 1 ISI record
1. Mathematics (Scienze matematiche ed informatiche)	72,06
2. Physics (Scienze fisiche)	85,51
3. Chemistry (Scienze chimiche)	89,93
4. Earth sciences (Scienze della terra)	67,66
5. Biology (Scienze biologiche)	83,4
6. Medicine (Sceinze mediche)	76,36
7. Agriculture and veterinary science (Scienze agrarie e Veterinarie)	64,01
8. Engineering and Architecture (Ingegnieria civile ed Architettura)	30,59
9. Industrial Engineering (Ingegneria industriale e dell'informazione)	76,05
10. Humanities (Scienze antichità, filologico-letterarie e storico-artistiche)	18,23
11. History, Phylosophy and Psycology (Scienze storiche, filosofiche, pegagogiche, psicologiche)	28,1
12. Law (scienze giuridiche)	12,25
13. Economics and Statistics (Scienze economiche e statistiche)	32,69
14. Sociology and Political Sciences (scienze politiche e sociali)	21,9
Missing**	14,39
Total	37,95

Table 3 - ISI WOS publications by reasearch areas (aree CUN)

3.3 National versus local competitions

In this framework we study the impact of the change in recruiting procedures occurred in 2000 for selecting (tenured) university professors in the Italian university onto the quality of selected professors. We study the quality of the selections under two alternative systems: national competitions, employed until the year 2000 (it consisted of two waves for associate professorship concluded in 1992 and 1998 and one wave for full professorship which ended in 1995 with some ties), and local competitions, held twice a year and granting three (later two) qualifications (idoneità) for each competition. This change was required by the explosion of university applications, combined with the reform of teaching curricula under the Bologna process. The national competitions were intended to be held biannually, but the actual occurrence was every 5-6 years.

In a nutshell the history of recruitment mechanism in Italy could be summarized as follows. Since 1979, standardized competitions were held to hire assistants, associate and full professors, and until 1998, almost all academic recruitment was substantially centralized. Despite the legislative prescription of one 'concorso' every two years, a three to four years interval occurred. National commissions of five members were chosen by lot within a pool of elected professors (from a pool of 15[8]) belonging to the same discipline. Commissioners declared which of the candidates had the qualifications to be promoted to associate/full professorship. Eligibility was given to a number of candidates greater then the available positions (usually 20% higher) for each discipline. Universities with opening positions drew by multilateral bargaining between them from the list of eligible applicants to fulfil their vacancies. Starting with 1999, recruitment procedures became entirely local, and each university could hold its own selection procedure (both for assistants, associates and full professors). Local commissions were comprised of five members: one belonging to the institution itself -the 'internal commissioner'- and the four others elected by the full set of Italian professors of that discipline. After 2005, a new reform act[9] established that the commission's members had to be drawn by lot in a pool of professors of three times the size of the local commission, elected by popular vote amongst the discipline's affiliates -of triple in size- elected by the whole set of discipline's affiliates. The commissions initially declared three qualified candidates for each 'concorso', but moved to two between 2007 until 2008, and only one thereafter. In the following years, universities with open vacancies could hire any candidate who had obtained a qualification. Professors hired under the new policy mechanism were engaged beginning in 2000, two years after the enactment of Berlinguer's reform[10]. Consequently, our empirical analysis marks the beginning of decentralization that year [11]. Unfortunately we do not have detailed information on all competitions, either under the national or the local system. We just observe changes in the level of appointment, that cannot be but the result of a public competition, since public universities are prevented from hiring or firing at will[12]. Nevertheless, the most appropriate definition would be promotions under a national/local competition systems. The number of such

[8]d.l n° 31/1979 and dpr n° 382/1980

[9]"Moratti reform", dl 230/2005

[10]dpr n° 390/1998 and dpr n° 117/2000

[11]this paragraph overlaps with Chapter 4, section 4.2.

[12]An exception is represented by appointment of professors from foreign universities, which did not require participation to a national or a local competition

promotions is reported in table 4. The most sizeable groups are represented by the transitions from assistant to associate professor and from associate to full professor, and we will mostly focus on them[13]

Year	Assist to Assoc	Out to Assoc.	Assoc to Full	Assist to Full	Out to Full	All Transitions
1991	80	46	137	11	21	295
1992	2039	749	51	1	15	2855
1993	237	180	58	3	23	501
1994	84	60	1285	62	77	1568
1995	30	569	175	21	539	1334
1996	35	142	49	6	118	350
1997	53	30	26	16	41	166
1998	2291	489	40	3	30	2853
1999	443	142	228	4	19	836
2000	1578	785	2285	70	429	5147
2001	2610	1288	2069	72	921	6960
2002	1914	803	1457	30	327	4531
2003	384	130	366	3	88	971
2004	686	317	493	6	159	1661
2005	2129	946	1555	26	258	4914
2006	1279	515	1023	10	212	3039
2007	378	227	319	4	143	1071
2008	174	161	154	2	83	574
2009	26	105	27	1	96	255
2010	573	159159	255	3	75	1065
2011	916	244	606	11	195	1972
Total	17939	8087	12658	365	3869	42918

Table 4 - Promotions of Italian professors

Despite being selected under different selection procedures, either national or local, any qualified candidate has to be selected by a local school/department in order to be hired. As a consequence, our analysis provides information on the quality of the recruitment in the Italian universities over the last two decades.

[13]There is an additional reason to leave out of our analysis the transitions from outside to professorship in an Italian university: since these candidates were previously working in a non Italian university, we do not match them to any previous publication. In addition, we ignore the size of the pool of potential applicants. Finally, this channel of recruiting received direct funding from the Ministry of Education of varying size over the years, following a request of local university and a check of subsistence of requirement of *clara fama*.

A final caveat concerns the number of promotions available. A quick inspection of table 4 reveals that more openings were available in the second sub-period (local system) when compared to the first one (national system).

However, what matters for a promotion is the relative odd. In figures 1 and 2 we show the ex-ante probability of being promoted, taking the ratio between the type of transitions (given in table 4) and the number of potential applicants (given by the stock of starting number of professors, given in table 1).

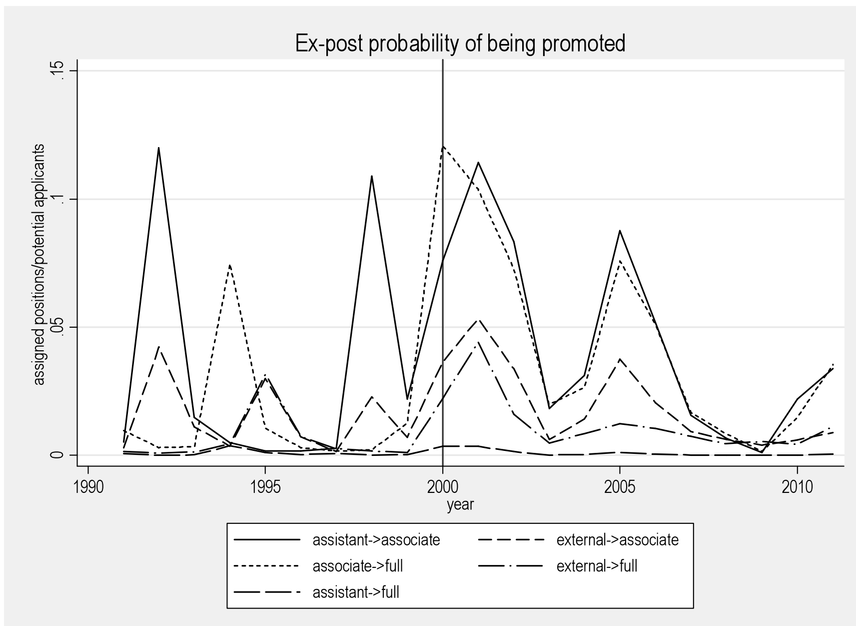


Figure 1 – Promotions over potential applicants, by type of transitions

Again by visual inspection, one may ascertain that the odds of being promoted are comparable across the two subperiods.

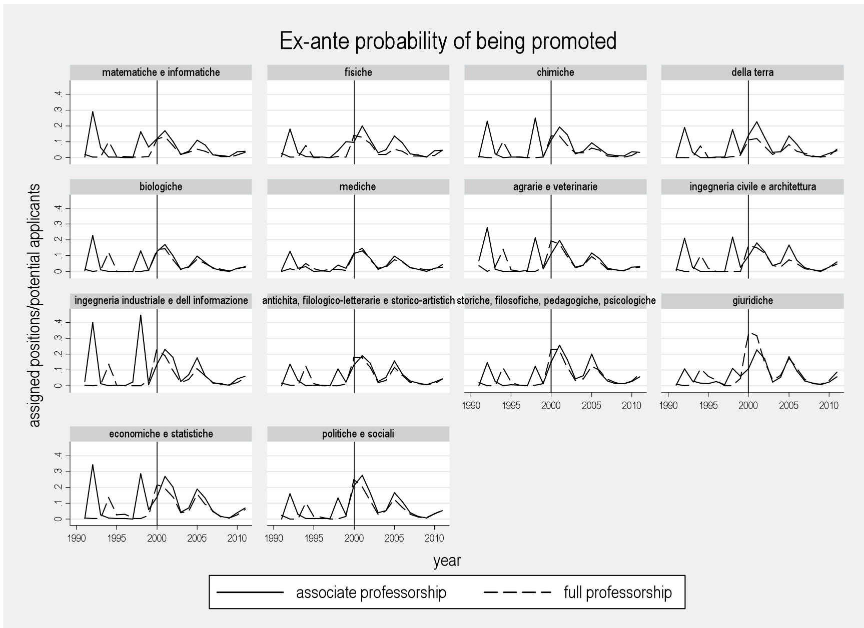


Figure 2 – Promotions over potential applicants, by research area

Summing up, in the sequel we will analyse the research productivity (as measured by WOS ISI records) of promoted professors in several research area (area CUN 1-7 and 9 – using most recent bureaucratic language, one could speak of "bibliometric areas"), comparing those promoted before and after the year 2000. This comparison provides an evaluation of the two selection procedure, national versus local.

3.4 The average quality of the promoted candidates

We start by providing some descriptive evidences of the distribution of the available measures of research outputs. By looking at figures 3, 4 and 5, we observe that researchers promoted to professors are more productive in the second subperiod, when we consider their number of ISI record[14]. However if we take their notoriousness (measured by the number of citations received by their articles) and/or their impact factor (measured by the average impact factor of the articles published that year) we do not observe significant changes over the two periods.

That our data on appearing in WOS are clearly trended is shown in table 5. If we estimate a linear probability model of the type:

$$y_{it} = \alpha_1 + \alpha_2 trend + \alpha_3 reform + \alpha_4 trend * reform + \varepsilon_{it} \tag{8}$$

were when $y_{it} = 1$ professor i has at least one ISI record, we see that the number of professors satisfying this condition is increasing (the coefficients of *trend* is positive), but at a lower rate (the variable *reform* contains a step-dummy assuming a unitary value starting with the year 2000).

Model	1	2	3
Variable	Assistants	Associates	Full Prof.
reform	-0.06554	-0.04105	-0.03064
	[0.00261]***	[0.00284]***	[0.00314]***
trend	0.02461	0.0284	0.02862
	[0.00039]***	[0.00042]***	[0.00048]***
reform*trend	-0,00911	-0.01328	-0.01416
	[0.00045]***	[0.00049]***	[0.00055]***
Obs.	434.694	363.405	328.155
R-squared	0.37	0.35	0.34

[14]Since the distribution of our measures of scientific productivity are rightly skewed, we consider their logs in order to show that they tend to a log-normal distribution. By so doing, we leave out of the graphs the candidates with zero publication records in WOS. Since we know that presence in WOS is increasing over the years, figures 3 to 5 provide a lower bound estimate of the distance between the two distributions.

Table 5 - Probability of having at least one ISI product, conditional on level of appointment - linear probability model; * significant at 10%; ** significant at 5%; *** significant at 1% robust standard errors in brackets - area and region controls included

We clearly need to identify a reference level, in order to detrend our measures. If we take the pool of applicants as our benchmark, we can compute the following measure of the quality of the promotions

$$q_{ijt} = \frac{y_{ijt}^{promoted}}{\frac{1}{n}\sum_{k=1}^{n_j} y_{kjt}^{non-promoted}} \tag{9}$$

where q_{ijt} is the relative quality of candidate i promoted in research area j at time t, computed as the ratio of individual publications record over the (equivalent) average record of all researchers in the same research area who are not promoted in the same year[15]. Thus all assistant professors represent the counterfactual for promoted associate professors, and similarly do associate professors for full ones.

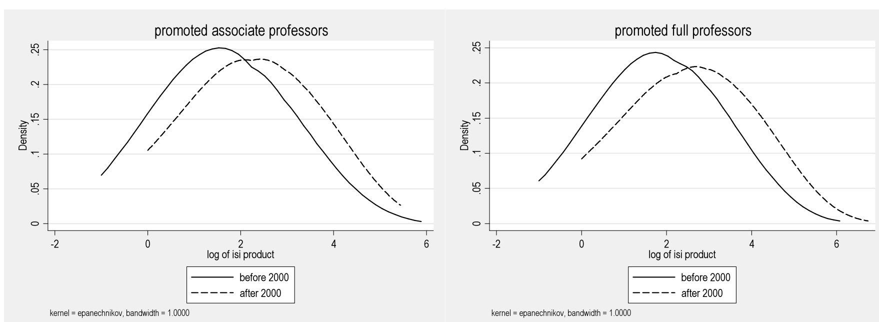


Figure 3 - Scientific productivity of promoted associate professors (3.a) /promoted full professors (3.b)

[15]By so doing we are implicitly assuming that whenever there is an opening for associate professorship, all assistant professors are potentially applying to the competition (and similarly with associate professors in case of competitions for full professorships). By anecdotal evidence, this assumption is realistic in the case of national competitions (when selection took place in a very irregular timing), while it is an approximation for the second subperiod (when applicants were entitled to apply to a maximum of five competitions per year. However, since each competition gave origin to three/two promotions, we think that this approximation may be acceptable.

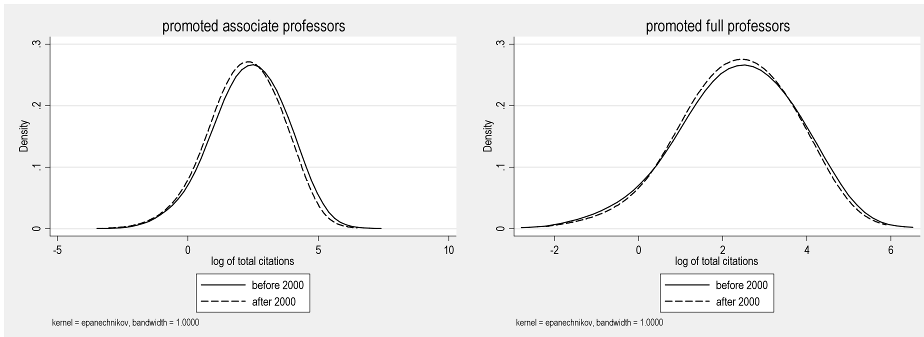


Figure 4 - Citations received by promoted associate professors (4.a) and promoted full professors (4.b)

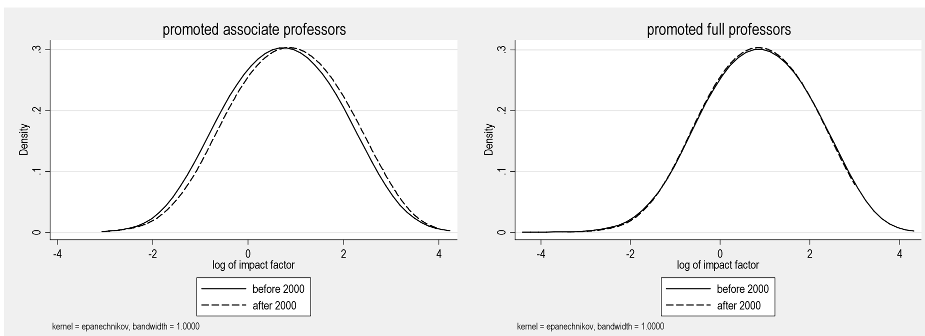


Figure 5 - Average impact factor of promoted associates (fig 5.a) and promoted full professors (fig 5.b)

If we replicate previous strategy we estimate

$$q_{ijt} = \alpha_0 + \alpha_1 trend + \alpha_2 reform + \alpha_3 trend * reform + \varepsilon_{ijt}, j = 1, 2, ..., 7, 9 \quad (10)$$

Results are reported in table 6 for associate professorships and in table 7 for full professorship. They indicate that there is some evidence of a decline in the quality of the selection associated to local competitions: when significant the intercept and the slope of these regression lines are lower in the second subperiod (this being more evident in the case of full professorships). A certain degree of heterogeneity is evident between academic disciplines. An overall -poorly significant- negative effect of reform (and no effect of the slope) onto bibliometric disciplines is associated with both associate and full professorship promotions. Reform effects are detected in Chemistry, Biology and Industrial Engineering for full professors selection procedures and in Physics, Chemistry and Biology for associate ones.

Variables/Area	all	1-7+9	1	2	3	4	5	6	7	9
trend	0.018	0.007	-0.082	0.097	0.09	-0.152	0.101	0.067	-0.081	0.025
	[0.049]	[0.022]	[0.062]	[0.041]**	[0.031]**	[0.198]	[0.045]**	[0.062]	[0.081]	[0.046]
reform	-0.398	-0.205	-0.289	-0.113	-0.354	-0.052	-0.705	-0.197	-0.127	-0.287
	[0.332]	[0.114]*	[0.337]	[0.264]	[0.199]*	[0.624]	[0.290]**	[0.425]	[0.366]	[0.221]
reform*trend	-0.059	-0.026	0.063	-0.152	-0.092	0.138	-0.102	-0.096	0.051	-0.044
	[0.053]	[0.024]	[0.067]	[0.047]***	[0.035]***	[0.207]	[0.048]**	[0.066]	[0.088]	[0.049]
Obs.	17.570	9.298	964	561	948	376	1.268	2.204	1.106	1.871
R-squared	0	0.03	0.06	0.06	0.07	0.036	0.07	0.07	0.05	0.03

Table 6 - Quality of the selection of associate professors, by research area – productivity measured by number of ISI products; *significant at 10%; ** significant at 5%; *** significant at 1% robust standard errors in brackets - region controls included

Variables/Area	all	1-7+9	1	2	3	4	5	6	7	9
trend	0.039	0.085	0.039	-0.022	0.394	-0.132	0.134	0.083	-0.111	0.178
	[0.131]	[0.052]	[0.131]	[0.309]	[0.110]	[0.295]	[0.100]	[0.114]	[0.081]	[0.084]**
reform	-0.433	-0.603	-0.433	-0.185	-2.649	0.095	-0.677	-0.48	-0.355	-0.982
	[0.767]	[0.331]*	[0.767]	[1.930]	[0.673]***	[1.573]	[0.613]	[0.743]	[0.532]	[0.471]**
reform*trend	-0.005	-0.11	-0.005	0.055	-0.399	0.117	-0.201	-0.106	-0.162	-0.187
	[0.133]	[0.053]	[0.133]	[0.310]	[0.111]***	[0.305]	[0.102]*	[0.115]	[0.097]*	[0.086]**
Obs.	565	6.592	565	416	655	235	905	1884	756	1176
R-squared	0.06	0.03	0.06	0.09	0.12	0.09	0.09	0.06	0.05	0.06

Table 7 - Quality of the selection of full professors, by research area – productivity measured by number of ISI products*significant at 10%; ** significant at 5%; *** significant at 1% robust standard errors in brackets - region controls included

Even if the first moment may have not changed, it is possible that the second (or higher moments) have changed. In figures 6 to 8 we plot the coefficients of variation computed for promoted professors over the three dimensions of scientific productivity we have available (number of ISI product, citations and impact factor). In this case we show that there is evidence of increased variability in the productivity of promoted associate professors, while the opposite trend reveals for full professorships.

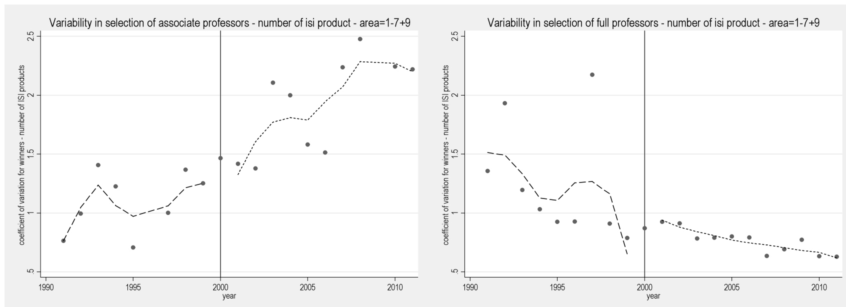


Figure 6 - Variability in selection of promoted associate professors – ISI product (fig. 6.a) and promoted associate professors (fig. 6.b)

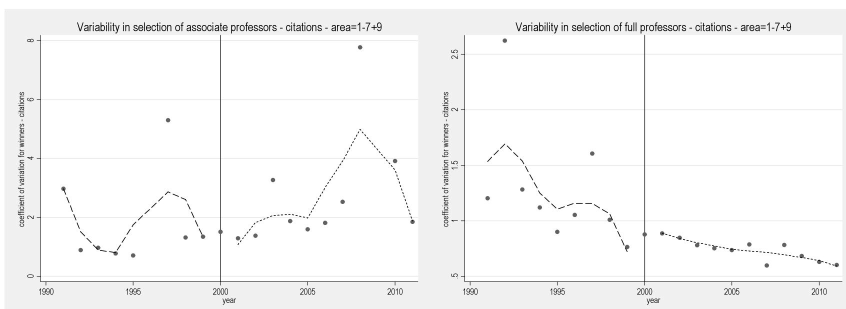


Figure 7 - Variability in selection of promoted associate professors - citations (fig. 7.a) and promoted full professors (fig. 7.b)

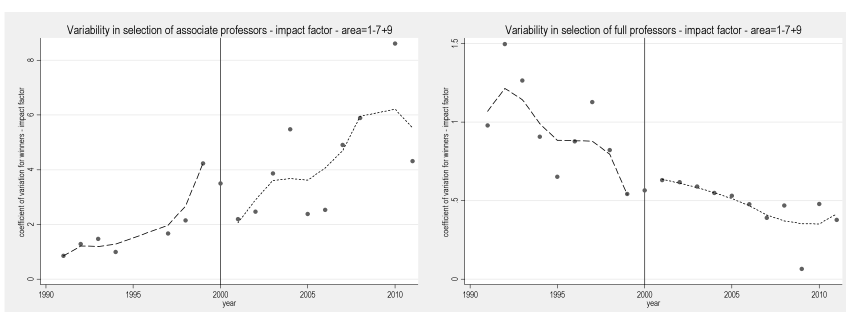


Figure 8 - Variability in selection of promoted associate professors – impact factor (fig. 8.a) and promoted full professors (fig 8.b)

3.5 The selection procedure

We now move deeper in the analysis of the selection process. As starting point, we proceed to counting the fraction of mis-allocated candidates, i.e. in a competition with k openings, we compute the fraction of non-winner with a rank (in terms of scientific productivity measure by ISI product) higher than k [16]. This measure is now computed at the level of each 372 research subfield, since it is at that level that ranking of candidates can be appropriately defined. The problem comes in with the local competitions, where we do not observe the participants to each local competition. Thus we have to interpret the reform as a regime change, where the number of vacancies is the sum of the vacancies available that year[17]In the following figures 9 and 10 we show the evolution of this measure of quality: there is a clear increase in this fraction overall and for each of the selected bibliometric disciplines, which on average increased over the whole sample from 45% to 54% for associate professorships and from 46% to 52% for full professorships.

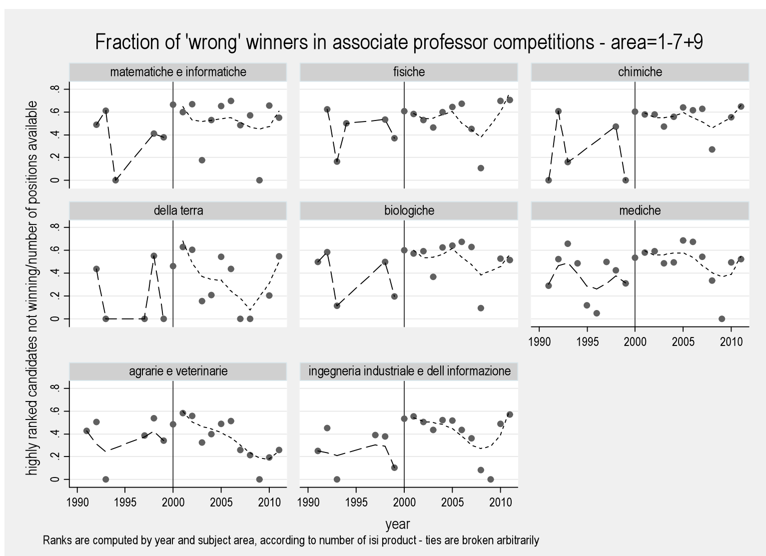


Figure 9 - Quality of selection by fraction of wrongly ranked associate professors winners, restricted to the selected disciplinary areas

[16]We thank Erich Battistin and Enrico Rettore for suggesting us this strategy

[17]There is another problem in computing such an index, which are ties. Not breaking the ties lead to indexes exceeding one, because there are long queues of zero productivity researchers. For this reason, we have chosen a procedure that arbitrarily break the ties.

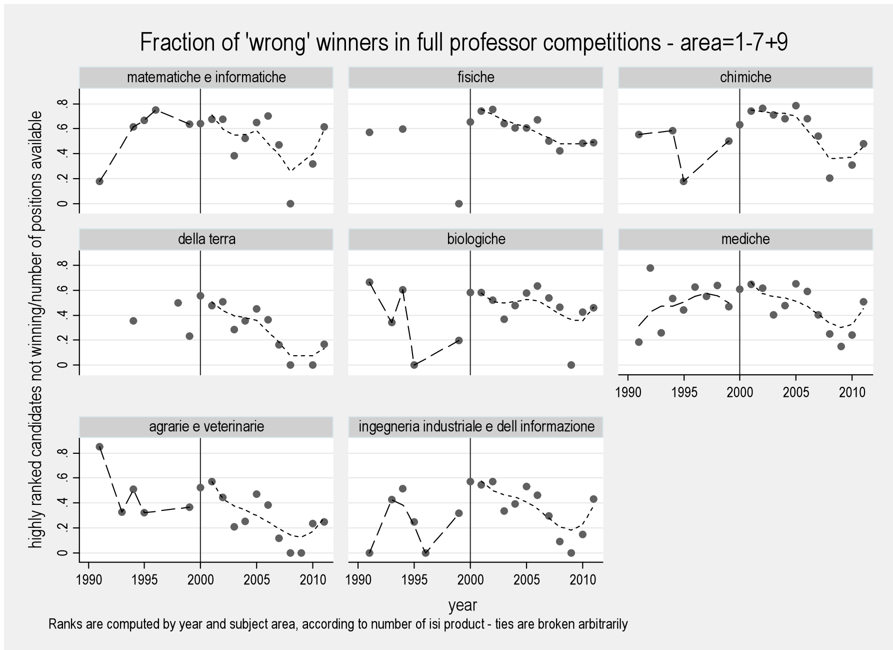


Figure 10 - Quality of selection by fraction of wrongly ranked full professors winners, restricted to the selected disciplinary areas

If we change our unit of analysis by collapsing the data by year and research subfield, we can compare the means before and after the reform (still using a subfield fixed effect control). In such a case we observe an increase in the index (more for associate than for full professors– see table 8). Thus the selection of professors under the local system seems to rely less on scientific productivity as measured by the number of ISI records. This is consistent with the increased variability in the quality of appointed professors, signalled by figures 6 to 8.

	1	2	3	4
	Associate Prof.	**Full Prof.**	**Associates 1-7+9 Area**	**Full 1-7+9 Area**
reform	0.078	0.071	0.079	0.032
	[0.011]***	[0.014]***	[0.015]***	[0.020]
ties	0.241	0.283	0.169	0.222
	[0.012]***	[0.011]***	[0.016]***	[0.020]***
Observations	3.492	2.952	1.829	1.534
R-squared	0.40	0.37	0.32	0.30

Table 8 -Quality of the selection of full professors, by research area - productivity measured by number of ISI products; *** significant at 1%, std, errors clustered by scientific sector in brackets - constant, year and subfield fixed effects included

The lower quality of promoted professors and the increased share of "wrongly" selected point in the direction of a change in the selection criteria, as a reflection of the change in the selecting committees. In order to investigate such a change, we estimate linear probability models for being promoted using all information available on the scientific productivity of the candidates (see table 9). As in the previous case, we take all the professors who are appointed at the inferior level as potential competitors for promotions. From this table we observe that scientific productivity (in terms of both quantity – ISI product – and quality – impact factor) and notoriousness (being cited) affect the probability of being promoted, more in the case of competition for associate professorships than in the case of competitions for full ones. The local competition seems to have shifted attention of the selecting committees from productivity to visibility.

Then a linear probability model and marginal effects of the probit model are estimated to study the effect of decentralization on the determinants of the probability of being selected as associate or full professors, conditional on being at previous stage (we consider standard career paths such as assistant promoted into associates and associates promoted to full professors as well as unusual carrers such as professors directly recruited into associates without being assistants before or assistants promoted directly to full professors).

Results (table 9) show evidences of positive effects of quantity and impact measures and negative effects of notoriousness for locally recruited on the probability of being promoted both to associates and to full professors with standard careers. Unclear results are associated with unusual careers.

from assistant to associate professors				
	1	**2**	**3**	**4**
n° of ISI records	0.000314			0.00316
	[0.00008]***			[0.00009]***
citations		0.00047		0.0002
		[0.00003]***		[0.00002]***
impact factor			0.00362	-0.00084
			[0.00017]***	[0.00018]***
reform	0.08354	0.0629	0.067	0.08177
	[0.00670]***	[0.00635]***	[0.00637]***	[0.00669]***
trend	-0.00472	-0.00268	-0.00269	-0.00491
	[0.00024]***	[0.00023]***	[0.00023]***	[0.00025]***
reform*trend	-0.00214	-0.00336	-0.00383	-0.00194
	[0.00081]***	[0.00074]***	[0.00075]***	[0.00081]**
reform*ISI products	-0.00019			-0.0002
	[0.00001]***			[0.00001]***
reform*citations		0.00004		0.00001
		[0.00001]***		[0.00000]***
reform*impact factor			0.00002	0.00011
			[0.00002]	[0.00002]***
Observations	253.465	253.465	253.465	253.465
R-squared	0.06	0.05	0.05	0.06

Table 9a - Probability of being selected as associates, conditional on being assistant professors, research areas 1-7+9, * significant at 10%; ** significant at 5%; *** significant at 1% robust standard errors in brackets - area and region controls included

from associate to full professorship				
	5	**6**	**7**	**8**
n° of ISI records	0.0014			0.0014
	[0.00005]***			[0.00005]***
citations		0.00045		0.00018
		[0.00003]***		[0.00003]***
impact factor			0.00256	-0.00087
			[0.00016]***	[0.00018]***
reform	0.06818	0.06545	0.06791	0.06633
	[0.00880]***	[0.00861]***	[0.00862]***	[0.00880]***
trend	-0.00329	-0.00199	-0.00195	-0.00338
	[0.00025]***	[0.00024]***	[0.00024]***	[0.00025]***
reform*trend	-0.00325	-0.00494	-0.00534	-0.00304
	[0.00105]***	[0.00099]***	[0.00100]***	[0.00105]**
reform*ISI products	-0.00009			-0.00009
	[0.00001]***			[0.00001]***
reform*citations		0.00003		0.00002
		[0.00000]***		[0.00000]***
reform*impact factor			0.00007	0.
			[0.00002]***	[0.00003]
Observations	212.828	212.828	212.828	212.828
R-squared	0.05	0.04	0.04	0.05

Table 9a - Probability of being selected as full professor, conditional on being associate, research areas 1-7+9, * significant at 10%; ** significant at 5%; *** significant at 1% robust standard errors in brackets - area and region controls included

Finally we may ask whether the increased variability in the scientific productivity of promoted professors may be the results of increased polarisation in the criteria followed by selecting committees (more correctly: by the departments who hired these professors). We have proxied this effect by computing the average indicator by year, university and research subfield (372 "settori scientifico-disciplinari") and we have interacted it with the individual measure of scientific productivity. We have also added a triple interaction with the reform in order to see whether something change after the reform. Each indicator of scientific productivity/visibility/impact is then interacted with the mean of the same indicator at university/research subfield[18]. These indicators behave in a strange way. Taken at face value (but notice the very small magnitude of the probability contribution) a candidate with a higher scientific productivity (number of Isi products) is less likely to be promoted/hired by a department/university with a higher (average) productivity (again measured by the number of isi products), but this effect is attenuated after the reform (implying that the reform favoured polarisation of behaviours). The opposite situation would occur when considering the other two indicators.

	1	2	3	4	5	6
	# isi product		citations		avg impact factor	
	assist to ass.	ass. to full	assist to ass.	ass. to full	assist to ass.	ass. to full
output	0.00431	0.00188	0.000116	0.00016	0.00146	0.00078
	[0.00011]***	[0.00007]***	[0.00003]***	[0.00003]***	[0.00017]***	[0.00017]***
reform	0.047	0.03815	0.05246	0.05255	0.06076	0.06098
	[0.0067]***	[0.00882]***	[0.00657]***	[0.00870]***	[0.00656]***	[0.00871]***
reform*time	-0.00632	-0.0056	-0.00728	-0.00751	-0.00807	-0.00813
	[0.00077]***	[0.00102]***	[0.00074]***	[0.00099]***	[0.00075]***	[0.00099]***
reform*output	-0.00023	-0.00013	0	-0.00001	-0.00018	-0.00019
	[0.00001]***	[0.00001]***	[0.00001]	[0.00000]**	[0.00002]***	[0.00003]***
avg out*indiv out	-0.00009	-0.00005	0.00008	0.00002	0.00032	0.00011
	[0.00001]***	[0.00000]***	[0.00001]***	[0.00000]***	[0.00003]***	[0.00002]***
reform*avg out*indiv out	0.00008	0.00004	0.00002	0.00003	0.00005	0.00006
	[0.00001]***	[0.00000]***	[0.00001]***	[0.00000]*	[0.00003]	[0.00002]***
Observations	253.465	212.828	253.465	212.828	253.465	212.828
Pseudo R-squared	0.07	0.05	0.06	0.104	0.05	0.04

Table 10 - probability of being selected as associate or full professors, conditional on being at previous stage; ** significant at 5%; *** significant at 1% robust std. err. in brackets - region controls included

[18]This is just a proxy of the quality of departments, since there may be more than one department in each university gathering professors of the same research subfield. We consider these people as members of the same "theoretical" department, since they should have been at least consulted during the hiring process of new professors.

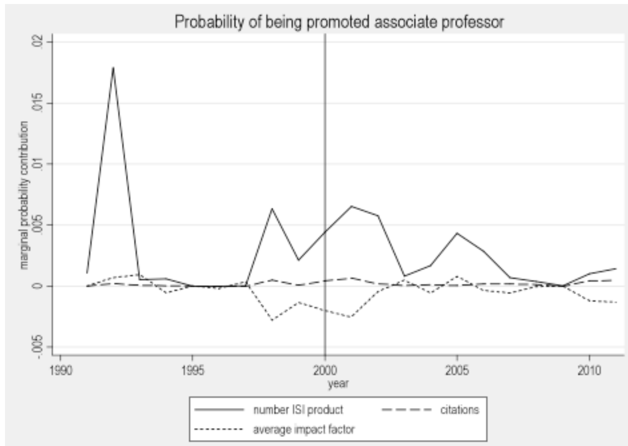


Figure 11a - Probability of being promoted from assistante to associate professor (marginal contribution in probability)

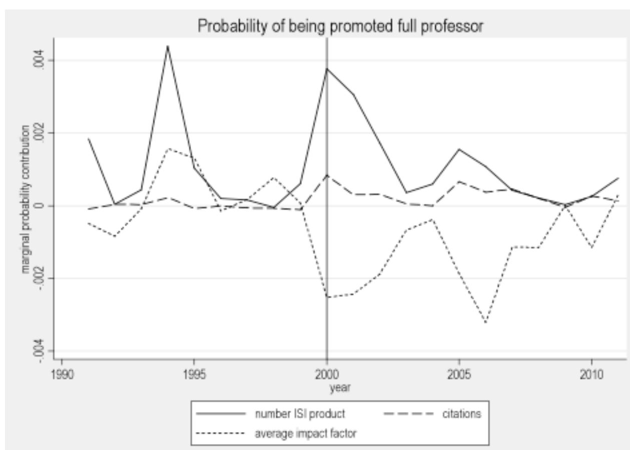


Figure 11b - Probability of being promoted from associate to full professorships (marginal contribution in probability)

3.6 Conclusions

The aim of this paper was to measure the decentralizing recruitment reform impact onto the quality of selected/promoted academic professors in Italy. Using a standard treatment approach we identify some evidence of a shrinking in the international research quality of researchers hired with local competitions (that means after 2000) in bibliometric diciplines only. An overall clear negative effects of the whole system is not significant.

A large degree of heterogeneity is evident between academic disciplines. In a nutshell results make evidence of an overall negative effect of reform (but no effect of slope) onto both associate and full professorship promotions in the bibliometric research fields. So that reform effects are detected in Chemistry, Biology and Industrial Engineering for full professors selection procedures and in Physics, Chemistry and Biology for associate ones. Moreover what is clear is the increase in the variability of the scientific productivity of locally promoted professors with respect to national ones.

Then going into details of the selection process we found an increase of the fraction of mis-allocated candidates (defined as the fraction of non-winners with higher rank with respect to the relative winners) at the level of each 372 research subfield in decentralized selection mechanisms (after 2000). This fraction of mis-allocated on average increased over the whole sample from 45% to 54% for associate professorships and from 46% to 52% for full professorships.

we explore the issue of the promotions criteria adopted by the selecting committees under the two regimes (national vs local). We found on one hand a higher importance of quantity and impact bibliometric measures for local selection committees (with respect to nationals) and on the other hand less importance of notoriety after decentralization.

Finally studying if the noticed increase in the variability of scientific productivity could be considered as the result of an increased polarisation in the criteria followed by selecting committees we found evidences that candidates with higher scientific productivity are unusually less likely to be promoted by a department with a higher productivity. But this particular effect is attenuated after the reform. So that we can argue that it inderctly implies that the reform has favoured polarisation of behaviours.

References

1. Allesina, S. (2011), “Measuring nepotism through shared last names: the case of Italian academia”, PLoS one 6(8), e21160;

2. Angelucci, M., G. De Giorgi, M. Rangel and Rasul I. (2010), “Family networks and school enrollment: evidence from a randomized social experiment”, Journal of Public Economics 94(3-4), 197–221;

3. Archambault E., Campbell D., Gingras Y., Larivire V. (2009), “Comparing Bibliometric Statistics Obtained from the Web of Science and Scopus”, Journal of the American Society for Informatin Science and Technology, 60(7);

4. Alberto Baccini A., (2010), “Valutare la ricerca scientifica”, il Mulino;

5. Braga M., Paccagnella M. e Pellizzari M, (2011), “Evaluating students' evaluations of professors”, Banca d'Italia Tema di discussione n. 825, ottobre 2011;

6. Checchi, D. (1999), “Tenure. an appraisal of a national selection process for associate professorship”, Giornale degli Economisti 58(2), 137–181;

7. Durante, R., Labartino G., Perotti R., and Tabellini G. (2009), “Academic dynasties”, Bocconi University Working Paper;

8. Salini, S. (2012), “An introduction to bibliometrics”, mimeo;

9. Fox, M. F. (1983), “Publication Productivity Among Scientists: A Critical Review”, Social Studies of Science, 13, 283-305;

10. Klavan and Boyak, (2007) “Is there a Convergent Structure of Science? A Comparison of Maps using the ISI and Scopus Databases1”, 11th International Conference of the International Society for Scientometrics and Informetrics meeting in Madrid, Spain, June 2007;

11. Labartino G., Perotti R., Durante R. (2011), “Academic Dynasties: Decentralization and Familism in the Italian Academia”, NBER Working Paper No. 17.572;

12. Levin S., Stephan P. (1991), “Research Productivity Over The Life Cycle: Evidence for Academic Scientists”, The American Economic Review, Vol. 81, Issue 1, 114-132;

13. Noser T., Manakyan H. and Tanner J. (1996), "Research Productivity and Perceived Teaching Effectiveness: A Survey of Economics Faculty", Research in Higher Education, Vol. 37 No. 3;

14. Seglen P.O. (1997), "Why the impact factor should not be used for evaluating research" British Journal of Medicine, 314, 498-50

4 Decentralized academic selection mechanisms: opportunity or parochialism?

Abstract

The aim of this paper is to test the impact of decentralizing academic selection mechanisms in Italy on scientists' research productivity following a policy reform introduced in 1998. Is decentralization an opportunity to select higher quality researchers or a way to increase parochialism? We focus primarily on the difference between individual research productivity (especially in terms of impact) before and after the reform, focussing on individual publication trends and differences in performance outcomes six-years after being hired. We collect ISI Web of Knowledge records with associated bibliometric indicators to obtain comparable international measures of research performance of the academics involved. We apply matching techniques (Propensity Score matching and the recently proposed Coarsened Exact Matching) to measure the impact of the policy change on individual research trends and on average research outcomes levels (measured six-year after selection) in a "quasi-experimental" research framework. Negative effects on research outcomes and trends of research impact measures are identified for people selected after decentralization controlling for disciplinary area and geographical effects. For research quantity indexes, the results show only slope differences.

JEL Classification: I23 C23 C52 Keywords: quasi-experimental design, policy evaluation, Coarsened Exact Matching, ISI WoK

4.1 Introduction

Has decentralization had a significant impact on the prevailing level of research productivity level of newly hired academics in Italy? To what extent does the mechanism of recruitment provide incentives for researchers to reach better publication scores after selection? The aim of our study is to determine whether local recruitment has an impact not only on the research quality of selected researchers but also on their mid-term research outcomes (examples of previous studies on research productivity among scientists are Fox, 1983 and Levin et al. 1991). We focus on the impact of the 1998 decentralization reform of the Italian university system on research outcomes for candidates publishing in international journals, bearing in mind that all other aspects of the system remained unchanged over the last two decades (salary benefits, university funding mechanisms etc...). We investigate this issue using datasets on Italian academics between 1991 and 2011. Data of ISI-Web of Knowledge (WoK) publications are collected to obtain standard and comparable bibliometric indicators of Italian professors' research activities while administrative records regarding affiliation, academic position and disciplinary area of professors are given by the National Ministry of Education, University and Research. We adopt a 'quasi-experimental' perspective applied to the Italian university system. Is it true that local recruiting performs poorly, and that Italian academics hired under the decentralized system are less productive? Although it is potentially easier to increase the discretional influence of local professors over the recruitment process with decentralization, has this opportunity made it simpler to match better candidates with local departments or has it merely increased the opportunity to engage in nepotistic behaviours? From a theoretical point of view, it is possible that decentralization (or parochialism) of recruitment mechanisms can reduce the incentives for candidates to produce international research outcomes (conference papers, journal articles etc..) and/or to submit papers to higher-quality scientific journals (which usually implies longer publication times and lower acceptance rates). More local management could generate the expectation that less stringent requirements will be applied. This consideration would be most crucial for applicants for assistant professorships and for assistant professors applying for full professorships. Our results document that decentralization, controlling for academic discipline, has a negative effect on ISI research outcomes (especially on measures of research impact) in mid-term performances of hired people.

4.2 The Italian Academic System

The Italian academic system is composed of 89 universities (28 private and 61 public) and 6 higher education institutions. The latter usually dispense only masters and PhD courses, being more research oriented than most universities. Three out of the 61 public universities are polytechnics. 11 out of the 28 private institutions are distance-learning universities. The university system is divided into 372 sectors of discipline (settore scientifico-disciplinare), grouped into fourteen research areas, as designated by the Italian National University

Council (CUN)[19]. Sectors of discipline are categorized for homogeneity within each research area, and the selection of research candidates is conducted by recruitment commissions within each academic discipline in both national and local recruitment systems. Considering academic disciplines as our reference level of analysis ensures validity in accounting for heterogeneity of recruitment behaviours between disciplines.

The Italian university system is constrained by national regulations. Each professor working at an Italian university is categorized by a level of arrangement (full professor, associate professor and assistant professor) and by one out of 372 sectors of discipline. Each vacancy is coded in a standardized format, and each filled position becomes tenured after a review conducted three years after hiring. Each position is also associated to a school (facoltà) for teaching duties and to a department for research activity. Salaries in public universities are set by law and vary only by level of arrangement and seniority. Schools and departments are prevented from differentiating wages among professors, linking payment to research productivity and/or teaching loads. As a consequence, in addition to celebrity and funds attraction, the strongest incentive to scientific productivity for individuals working in academia derives from expected promotion (being hired as assistant professor, being promoted associated or full professor).

Given the public nature of the employment contracts, university professors can only be hired through public competitions that should grant publicity of the vacancy, selection of the selecting committee based on objective criteria, transparency of the selection process. This may explain why it is crucial for the research productivity and quality to study the different incentives designed by different selection procedures. Reforms in 1998 changed these procedures with respect to several dimensions:

1) <u>level of selection</u> (national or local, which mostly affects the number of competing applicants, but also the timing of the selection due to the heavier bureaucratic load associated with a nationwide competition);

2) <u>selection of committees</u> (in accordance to the co-optation attitude of academia, for most of the period under analysis the committees were elected out of professors of the same sector of discipline, with element of randomness introduced at some stage);

[19] Mathematics and Computer Sciences, Physics, Chemistry, Natural Sciences, Biology, Medicine, Agriculture and Veterinary, Civil Engineering and Architecture, Industrial Engineering, Literature, History, Psychology, Law, Economics and Statistics, Social Sciences

3) <u>number of eligible applicants</u> (each ‘concorso’ declares a number of winners that are eligible to become professors, this number is usually equal to, but sometimes greater than, the number of available vacancies).

Since 1979, standardized competitions were held to hire assistants, associate and full professors, and until 1998, almost all academic recruitment was substantially centralized. Despite the legislative prescription of one ‘concorso’ every two years, a three to four years interval occurred. National commissions of five members were chosen by lot within a pool of elected professors (from a pool of 15[20]) belonging to the same discipline. Commissioners declared which of the candidates had the qualifications to be promoted to associate/full professorship. Eligibility was given to a number of candidates greater then the available positions (usually 20% higher) for each discipline. Universities with opening positions drew by multilateral bargaining between them from the list of eligible applicants to fulfil their vacancies (Checchi shows some evidences from a single national selection procedure for associate professorship, 1999).

Starting in1999, recruitment procedures became entirely local, and each university could hold its own selection procedure (both for assistants, associates and full professors). Local commissions were comprised of five members: one belonging to the institution itself -the ‘internal commissioner’- and the four others elected by the full set of Italian professors of that discipline. After 2005, a new reform act[21] established that the commission’s members had to be drawn by lot in a pool of professors of three times the size of the local commission, elected by popular vote amongst the discipline’s affiliates -of triple in size- elected by the whole set of discipline’s affiliates. The commissions initially declared three qualified candidates for each ‘concorso’, but moved to two between 2007 until 2008, and only one thereafter. In the following years, universities with open vacancies could hire any candidate who had obtained a qualification.

Professors hired under the new policy mechanism were engaged beginning in 2000, two years after the enactment of Berlinguer’s reform[22]. Consequently, our empirical analysis marks the beginning of decentralization that year.

[20]d.l n° 31/1979 and dpr n° 382/1980
[21]“Moratti reform”, dl 230/2005
[22]dpr n° 390/1998 and dpr n° 117/2000

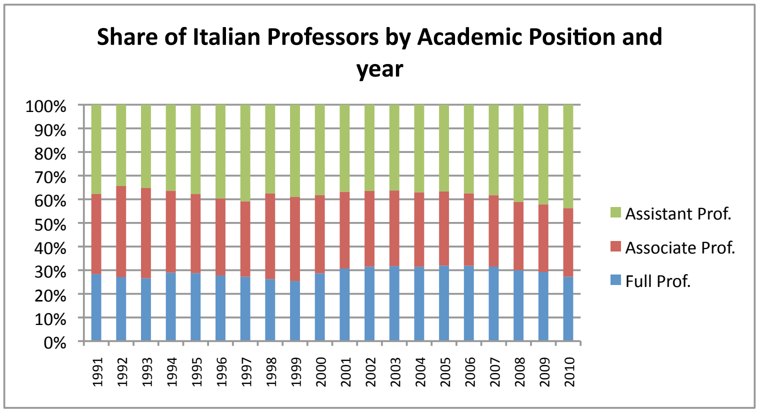


Figure 1: Share of Italian Professors by Academic Position and year

4.3 Theoretical incentives

The decentralization of academic recruitment was an exogenous shock to recruitment rules that potentially impacted the subsequent career of selected professors. The majority of other factors that affecting research performance over this time have remained constant, setting the stage for a natural experiment for considering the effect of decentralization. Notwithstanding the hiring changes, university' funding mechanism remained totally disconnected from managerial behaviours. No salary incentives (or penalties) were provided to incentivize (or prevent) virtuous behaviours of the commissioners in selecting high (or low) quality candidates. Student evaluations of teaching performance for new hired professors usually had no impact on the professional life of professors (although including aspects of these evaluations become compulsory after 2000). Due to strict privacy rules, the results of evaluation exercises were in most of the cases known only by each professor until recently.

No evaluation mechanisms were established to assess recruitment procedures at department (or university) level by the central administration. Thus, over the time period we consider, there was no private cost for opportunistic behaviour for (part of) the selecting committee, as well as no impact on institutional funding mechanisms, except for lower scientific reputation.

Decentralizing academic recruitment could have mixed effects. On the one hand, decentralization could improve productivity and efficiency for at least three reasons; first, local recruitment usually induces speedier selection procedures (national “concorsi” were held every 4-6 years), second, more certainty of fulfilling available vacancies when needs arise (both for research and teaching necessities), and third, decentralization could lead to more competition among universities to attract candidates. An additional may be better matching between academics and institutions.

In practice, the multilateral bargaining between winners and hosting universities could last one or two years under national mechanisms; secondly it means less compromise with “The Academia” when it is dominated by few national prevailing “Schools”. The Italian academia is certainly not so unfamiliar with such corporatism (Durante et al. 2009; Labartino et al. 2011; Allesina, 2011; Angelucci et al. 2010) and the existence of “schools” that could exert a direct influence on the selected candidates has to be considered as a possible problem. Last but not least it could enhance more competition among universities in attracting better candidates. The better candidate could also be the one that particularly fits with institutional needs (in terms of research competences and experiences).

On the other side, decentralized selections enable institution to favour individuals based on familiar, professional or political considerations, independent of their experience or qualifications. Local processes also may lead to less competition with respect to national procedures.

In addition, it is important to analyse the extent to which influence of the selection procedures induces behaviours of the potential candidates. Before reforms, national concorsi were held less frequently and involved a larger number of interested applicants. This meant more competitors. In most circumstances, a greater number of competitors encourage more effort on the part of the candidates, providing incentives for individuals to maximize their probability of winning through performance. Furthermore, the number of peers under central selections was stable in the Italian context, meaning that opportunities for advancement could be considered equal over time for each concorso.

Decentralizing academic recruitment meant fewer competitors participate in concorsi both in the present, and future. Geographical constraints are also important because, with a local system, there is an incentive to participate in concorsi that are relatively close to the candidate’s home area, rather than compete in all Italian concorsi for that year. If publishing more papers in impact journals increases an individual’s probability of being selected (as should theoretically be the case), decentralizing academic procedures could provide fewer incentives for local candidates. Indeed the individual choice of putting less or more effort is not mainly driven by the own candidate willingness to exert it but it could probably be given by the selection year: before and after 2000 is the threshold. Our research question is twofold; i) has decentralization of selection mechanisms improved (or worsened) mid-term research outcomes of winners? and, ii) have local recruitment mechanisms incentivized (or discouraged) individual research careers?

4.3.1 A simple theoretical model

The starting teorethical model we consider here originates from Lazear[23], on relative compensation. In a framework where promotion (and associated wage increase) is attributed to the best performer among two identical workers, and performance depends on both luck and effort, the worker will supply more effort the higher is the wage premium and the lower is the variance of the random component.

Basic version (deterministic)

Le t us define w_1 as the wage rate before promotion (i.e. associate professor wage rate – better thinking of the monetary value of the prestige connected to being an associate professor) and w_2 as the wage rate after promotion (full professor wage rate – the same model can also be thought with respect to the promotion of assistants to associate professorships). Then $p = \frac{m}{n}\pi\left(e\right), \pi' > 0, \pi'' < 0$ is the probability of obtaining the promotion. It is a positive function of effort e, and it exhibits decreasing marginal productivity. Then let n be the number of applicants and m the number of posted vacancies.

$c = c\left(e\right), c' > 0, c'' > 0$ is the monetary-equivalent cost of effort which exhibits increasing marginal cost. The monetary-equivalent cost of effort takes it's explicit functional forms as follows $c(e) = \frac{\delta e^2}{2}$ and also $\pi(e) = e^{\alpha}, \alpha > 1$.

Each applicant maximises the expected gain from participating to the competition

$$Max_e p(w_2 - w_1) - c$$

The th first order conditions are:

$\frac{m(w_2-w_1)}{n}\alpha e^{\alpha-1} - \delta e = 0 \Leftrightarrow e = (\frac{m}{n}\frac{\alpha(w_2-w_1)}{\delta})^{\frac{1}{2-\alpha}} = f\left(m_+, n_-, \alpha_+, \delta_-, w_2 - w_{1+}\right)$

[23]Chapter 3, Personnel economics, MIT 1995

The effort (that could represented by a proxy given by the scientific productivity) is increasing in the size of the premium and in the productivity of the effort α, while it is decreasing in the cost of effort δ. What is less satisfactory is the way in which competition is modelled, since an increase in participation n lower the incentives. Since local competitions are characterised by higher $\frac{m}{n}$, the model would predict *higher effort* as consequence of the reform (contrary to the empirical evidence).

So far all candidates are ex-ante identical. Without cost we can introduce agent heterogeneity in the cost function assuming that abler candidate have a lower cost of effort. Indicating with a_i the individual ability endowment, we can rewrite $c(e) = \frac{\delta e^2}{2}$ obtaining $e = (a_i \frac{m}{n} \frac{\alpha(w_2 - w_1)}{\delta})^{\frac{1}{2-\alpha}} = f\left(m_+, n_-, \alpha_+, \delta_-, w_2 - w_{1+}, a_{i+}\right)$ Better candidates put more effort and are more likely to obtain promotion.

Basic version (stochastic)

Suppose that individual scientific productivity y_i depends on effort and luck (being accepted by top rank journal partially depends on factors beyond individual control)

$y_i = e_i + \varepsilon_i$, where luck is uniformly distributed over the interval $[-b, +b]$[24].

Competition is governed by the following rule: promotion(s) is(are) allocated to best performer(s). The simplest case (see Lazear 1995 – see also P.Garibaldi, Personnel Economics in Imperfect Labour Markets, Oxford University Press 2006, chpt.8) is when there are only two participants (i and j) and one position, which is allocate to the best performer. In such a case each player maximises:

$Max_{e_i} pw_2 + (1-p)w_1 - \frac{\delta e_i^2}{2} \backsimeq p(w_2 - w_1) - \frac{\delta e_i^2}{2} = prob(y_i > y_j)(w_2 - w_1) - \frac{\delta e_i^2}{2} = prob(e_i + \varepsilon_i > e_j + \varepsilon_j)(w_2 - w_1) - \frac{\delta e_i^2}{2}$

Assuming that ε is iid, we can rewrite as follows:

$prob(e_i + \varepsilon_i > e_j + \varepsilon_j) = prob(e_i + e_i > \varepsilon_j + \varepsilon_j) = F\left(e_i - e_j\right) = \frac{(e_i - e_j) - b}{2b}$

Then taking FOC:

$\frac{dF(e_i - e_j)}{de_i}(w_2 - w_!) - \delta e_i = f(e_i - e_j)(w_2 - w_1) - \delta e_i = 0$

By symmetry (equivalent to a Nash equilibrium among two players) $e_i = e_j$ and therefore

$e_i = \frac{f(e_i - e_j)(w_2 - w_1)}{\delta} = \frac{f(0)(w_2 - w_1)}{\delta} = \frac{(w_2 - w_1)}{2\delta b}$ which means that the effort is increasing in the prize, decreasing in its cost and in the variance of the random component.

[24] This implies that $E[\varepsilon] =, \mathrm{var}(\varepsilon) = b^2/12, F(x) = prob(\varepsilon \leq x) = \frac{x-b}{2b}, f(x) = \frac{1}{2b}$.

If we consider now the case where we have n participants and one vacancy, in order to be promoted a candidate should cearly have the highest performance. This event has the following probability:

$prob(y_i > y_{-i}) = \prod_{j=1,j\neq i}^{n} prob(y_i > y_j) = \prod_{j=1,j\neq i}^{n} prob(e_i - e_j > \varepsilon_i - \varepsilon_j) = \prod_{j=1,j\neq i}^{n} F(e_j - e_i)$

Following the previous strategy as before, the FOC

$\frac{d[\prod_{j=1,j\neq i}^{n} F(e_j - e_i)]}{de_i}(w_2 - w_1) - \delta e_i = (w_2 - w_1)\sum_{j=1}^{n} f(e_i - e_j)[\prod_{k=1,k\neq i,k\neq j}^{n} F(e_i - e_k)] - \delta e_i = 0$

Using the symmetrical Nash equilibrium assumption

$(w_2 - w_1)\sum_{j=1}^{n} f(0)[\prod_{k=1,k\neq i,k\neq j}^{n} F(0)] - \delta e_i = (w_2 - w_1)\sum_{j=1}^{n} f(0)[\frac{1}{2^{n-1}}] - \delta e_i = (w_2 - w_1)\frac{n}{2b}[\frac{1}{2^{n-1}}] - \delta e_i = 0$

which yields $e_i = \frac{(w_2 - w_1)}{2\delta b}\frac{n}{2^{n-1}}$

This expression is nonlinear in the number of participants: an increase in participation to the competition for lower number may increase effort, but for large numbers effort is decreasing.

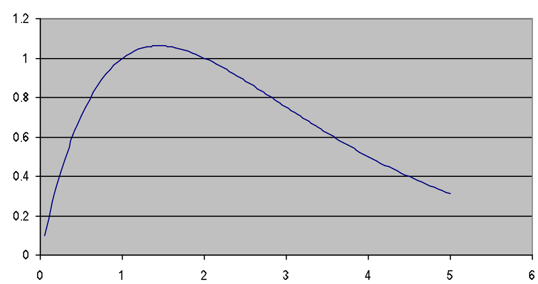


Figure 2 - Effort function representation

If we change the number of vacancies (m vacancies for n applicants) and we assume that a candidate is promoted as long as her performance exceeds the performance of at least $(n - m)$ competitors, the model is modified as follows

$prob(y_i > y_j), \vee j \neq i, j = 1, ..., (n-m) \Leftrightarrow \prod_{j=1,j\neq i}^{n-m} prob(y_i > y_j) = \prod_{j=1,j\neq i}^{n-m} \frac{e_i - e_j - b}{2b}$

and following identical steps we end up with

$e_i = \frac{(w_2 - w_1)}{2\delta b}\frac{n-m}{2^{n-m-1}}$

All these models point to the idea that effort is driven by *relative* probability of winning.

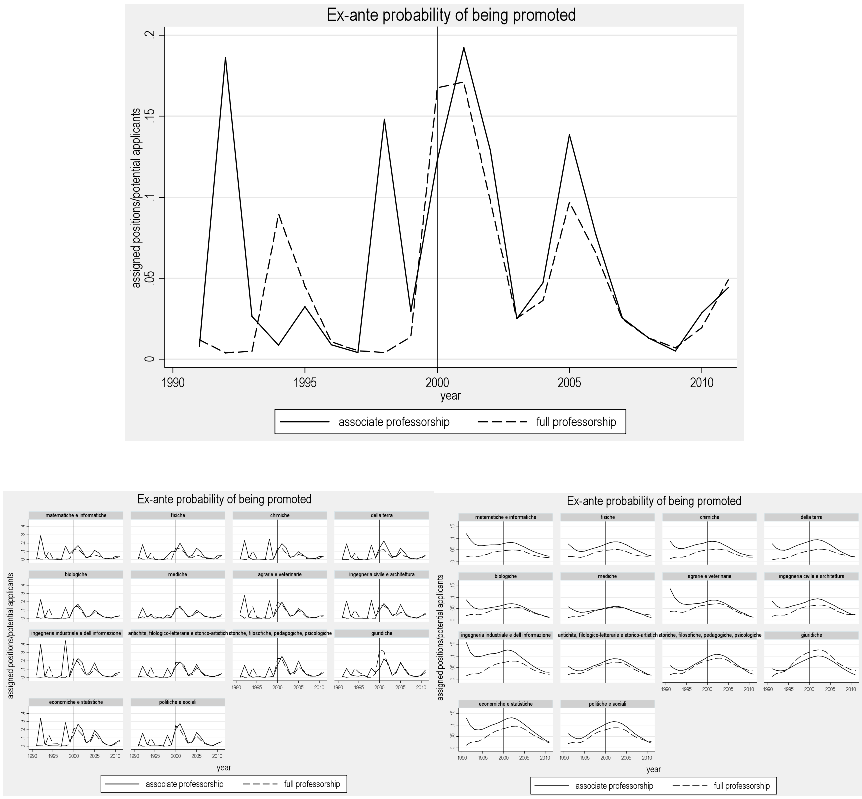


Figure 3 - Ex-ante probability of being promoted overall and by scientific disciplines for both associates and full professors

Thus we have computed the number of winners over the potential applicants in each year and for each level of competition. All of them are consistent with the idea that there have been a decline in relative probability, pulling down scientific productivity (thus the relative worsening would be the results of lower incentives created by the reduction in the number of qualifications (*"idoneità"*) and/or in the budget allocated to new hiring. In addition remember that newly promoted associate professors *ceteris paribus* reduce the relative probability of being appointed full professors.

4.4 Data

Data for this paper were collected from two primary sources: MIUR – Italian "Ministry of Education, University and Research" - and the "bibliographic" Web version of WoK database. We do not use alternative databases from Scopus and Google Scholar. To quantitatively determine an individual's production of international scientific research, we select ISI WoK as our primary information source.

Web of Knowledge, powered by the Institute for Scientific Information and distributed by Thomson Reuters, has been the standard in the bibliometric field for the past 30 years and indexes more than 8.700 journals in the fields of arts, humanities, sciences and social sciences. Scopus, published by Elsevier (www.info.scopus.com), indexes a greater number of journals (12.850, including 500 open access journals) within the medicine, technical and social sciences. Scopus is significantly larger in size and covers more of the international literature, but it excludes the humanities (Figure 4; Klavan and Boyak, 2007). Google Scholar stands-today-as the main potential competitor of ISI and Scopus (particularly in light of the fact that it is the only one without commercial interests), but currently has outstanding information reliability problems.

Using bibliometric databases for evaluation purposes is often debated (Falagas et al., 2008; Bakkalbasi et al., 2006). Despite significant disagreement concerning specific research questions, most points of view agree that not all sources are ideal for all circumstances, and the choice of database to use should take into account the aims of the research (De Battisti, F., Salini, S., 2011).

In this context, our choice in using ISI-Web of Knowledge is due representativeness requirements, in addition to our need for high accuracy of information. The availability of standard metrics information (e.g. Journal Citation Reports) was also a central consideration. ISI, while maintaining a clear excellent coverage in science, is less affected by lack of information in the fields of arts and humanities (AHCI) than Scopus. Scopus otherwise would be preferable as a greater number of indexed journals and for the best coverage in the social sciences (S/SS) (Norris, M. and Oppenheim, C., 2007).

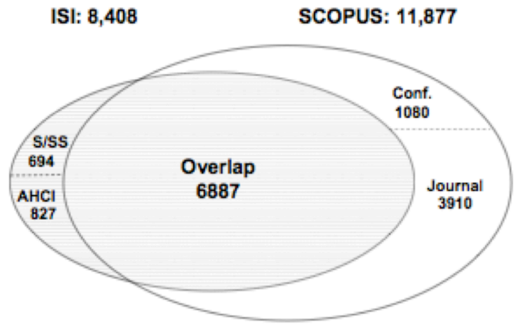


Figure 4: Overlapping and unique coverage of ISI and Scopus databases, 2004 (Klavans e Boyak, 2007)

The literature documents the presence of high correlations between and among bibliometric measures obtained by Scopus and ISI databases (Archambault et al., 2009). Thus, despite using a single source, we expected robust results of our analysis using few bibliometric measure. The main drawback we face still remains the disciplinary coverage specification of international research data with respect to the whole research domain.

Researchers of some disciplines such as History and Literature usually publish on national journals only (usually with articles written in Italian). For these academics, little bibliometric information is available on ISI, SCO or GS. Distortion of data due to higher 'ISI exposure' of some disciplines in comparison to others can only be managed by disaggregating the analysis by discipline, which is the approach we employ.

Information regarding academic positions, disciplinary areas, and university affiliation are available online from 2000 to 2011. We obtained data on academic careers before 2000 from Cineca, a MIUR agency which collects administrative data on personnel as well as on competition for professorship in Italy. These data have several known problems, often relating to the uniqueness of identifying codes of individuals, and missing data on academic disciplines over the first five years (1990-1995). After we corrected for these issues to the best of our ability, we found a 1.5% degree of imbalance with respect to the last available official statistics published by Miur (reported in table 1). These differences are likely are due to a few rare categories of professors, such as newly hired associate and full professors attracted from abroad, fixed-contract new researchers positions and so on. However, such a small difference is unlikely to seriously bias our results, or be the cause of distortionary effects in our estimation procedures.

Final database					Official statistics					%Δ
Year	Full Prof.	Assoc.	Assist.	Tot.	Year	Full Prof.	Assoc.	Assist.	Tot.	
1991	11750	14042	15642	41434	1991	nd	nd	nd	45248	0,08%
1992	11804	16746	14964	43514	1992	nd	nd	nd	nd	.
1993	11876	17084	15739	44699	1993	nd	nd	nd	47839	0,06%
1994	13288	15915	16694	45897	1994	12856	15046	20230	47824	0,04%
1995	14011	16313	18417	48741	1995	13204	15465	19891	49098	-0,01%
1996	13719	16093	19583	49395	1996	13720	16231	21038	48560	0,0%
1997	13399	15675	20105	49179	1997	13402	15619	20167	49187	0,0%
1998	13098	18108	18748	49954	1998	13103	18108	18745	49956	-0,2%
1999	12905	18069	19815	50789	1999	12899	17863	19949	50711	1,9%
2000	14411	16615	19200	50226	2000	14676	16973	19542	51191	-1,9%
2001	16901	17879	20255	55035	2001	16418	17572	20011	54001	-2,3%
2002	18148	18504	21055	57707	2002	17571	18100	20714	56385	-2,1%
2003	17997	18115	20577	56689	2003	17388	17783	20371	55542	-2,2%
2004	18062	18094	21341	57497	2004	17469	17633	21149	56251	-0,9%
2005	19296	18982	22186	60464	2005	19147	18849	21904	59900	-0,9%
2006	19843	19084	23355	62282	2006	19676	18966	23099	61741	-0,5%
2007	19640	18776	23793	62209	2007	19623	18739	23560	61922	-0,5%
2008	18929	18253	25923	63105	2008	18932	18261	25569	62762	-1,1%
2009	17980	17630	25911	61521	2009	17878	17567	25434	60879	-2,0%
2010	15949	16967	25590	58506	2010	15834	16745	24784	57363	0,0%

Table 1: Official 1998-2010 data reported in 11° rapporto CNVSU (tab 5.8 pp.154)

Individual and research output information refers to the period 1990-2011. Italian academics' publication records were downloaded using specific institution and publication year as key query parameters.

Our work provides evidence of some limitations with regard to ISI web interface, which makes bibliometric analyses more difficult: first, of all the constraint on the maximum number of articles (500) that can be downloaded for each query. In addition "homonymity" issues should be considered in assigning each product to it's own author. Furthermore a number of duplicate and incomplete records are detected and definitively deleted from the dataset.

After filtering, our data include around 1.000.000 ISI products over the last 20 years. Duplicates and incomplete records were deleted obtaining a consistent database of 963.181 scientific publications with at least one Italian author.

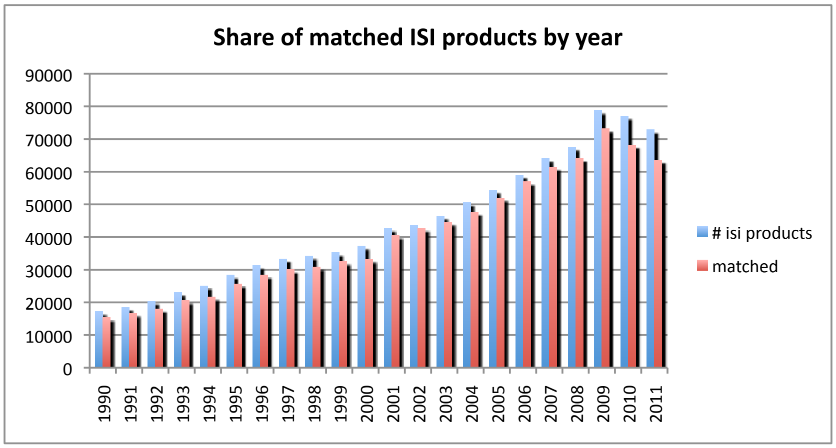


Figure 5: Share of ISI products with at least a 1:1 matching on author's identifying code by publication year

We then employ a three-step matching procedure to assign the corresponding author identifying codes to each research product (it might be possible that one paper is co-authored by two or more different individuals belonging to the Italian academy). In the first step, a combination of the institution code, publication year, last name and name initial letter as the linking-rule and a 60.5% of the entire number of ISI records is attributed this way to its own author(s) (excluding individuals with at least one homonymous author in the current year). The second step, performed on excluded papers, is running a matching algorithm based on the previous rule, augmented by control individuals with homonymous authors in the current year but not in the same institution. An increase of 30% over the initial record amount is attributed in this way. The third step consists of matching the left-overs with last name and name initial letter as new linking-rule, allowing the procedure to check for records stored under different institutional codes (considering the case of moving researchers between different universities). Overall, 91% of matched records could be considered as a good result of our procedure. Indeed a 9% percent of ISI products is plausibly stored in the database reporting an Italian affiliation but with an author who is not included in the official faculty list provided by MIUR, which may be plausible due to post-doctoral students, PhD candidates and individual researchers not included in MIUR's research and teaching faculty list. In addition to this, it is interesting to note that the ISI matching rate of Italian professors is strongly increasing in the last twenty years. Starting with a 14% rate of academics with at least on ISI publication record in 1991 we note an increasing growth rate over the two decades until the 60% of 2011 (Table 2).

Year	At least 1 ISI paper		Tot.
	No	Yes	
1991	86,03	13,97	100
1992	83,14	16,86	100
1993	80,94	19,06	100
1994	78,87	21,13	100
1995	76,35	23,65	100
1996	74,35	25,65	100
1997	72,51	27,49	100
1998	70,12	29,88	100
1999	68,15	31,85	100
2000	65,71	34,29	100
2001	62,24	37,76	100
2002	59,06	40,94	100
2003	57,19	42,81	100
2004	55,14	44,86	100
2005	52,07	47,93	100
2006	49,12	50,88	100
2007	46,84	53,16	100
2008	44,21	55,79	100
2009	42,32	57,68	100
2010	40,52	59,48	100
2011	38,21	61,75	100
Total	62,05	37,9	100

Table 2 - Percentage of professors with at least one paper on ISI by year

The overall percentage of academics with at least one paper on ISI over the entire period is higher for full professors with respect both to associate and assistant professors (50%). This is a measure biased by the lengthier time of exposure for full professors and associates.

Academic Position	At least one ISI paper (1991-2011)		
	No	Yes	Tot.
Assistant Professor	50,98	49,02	100
Associate Professor	45,17	54,83	100
Full Professor	41,41	58,59	100
Total	46,32	53,68	100

Table 3 - Percentage of professors with at least one paper on ISI by Academic Position

Our data show different numbers of individuals over time, the panel for each academic professor is unbalanced due to varying entry points into the administrative archives of MIUR. Data vary depending on the year of selection, the year of the first published international paper, the persistence rate of publication on ISI of the own discipline. Discontinuities are also possible (and several are identified in our analysis), due to such considerations as working abroad, or that the individual enters unusually one year and ceases publishing after that point. These issues could have significant effects on the distribution of bibliometric indicators over time, and the challenges are more pronounced in some disciplines than in others. We arbitrarily decide a threshold in order to exclude disciplines with lower level of individuals with at least one product on ISI. Heterogeneity within academic research areas is highly effective in ISI studies, and the following table (Table 4) give us the idea of which of the standard disciplines overcome the 50% cut-off level of individuals with at least one record. We consider in this study especially the scientific area where this percentage is greater than 50% with respect to the different historical and individual nature of each discipline. Bibliometric indicators could be theoretically considered for all the scientific areas but a rate grater than 50% guarantee a degree of reliability in our research exercise. A high sensitivity level among academic fields and aggregation levels is common in bibliometric studies.

Thus we reduce the area on less heterogeneous disciplines this way and build the dataset longitudinally, considering year-of-selection as the (moving) starting point for all individuals considered.

Disciplinary Area	id	At least 1 ISI paper		Tot.
		Yes	No	
Mathematics and Computer Sciences	1	27,94	72,06	100
Physics	2	14,49	85,51	100
Chemistry	3	10,07	89,93	100
Natural Sciences	4	32,34	67,66	100
Biology	5	16,6	83,4	100
Medicine	6	23,64	76,36	100
Agriculture and Veterinary	7	35,99	64,01	100
Civil Engineering and Architecture	8	69,41	30,59	100
Industrial Engineering	9	23,95	76,05	100
Humanities	10	81,77	18,23	100
History, Phylosophy and Psycology	11	71,9	28,1	100
Law	12	87,75	12,25	100
Economics and Statistics	13	67,31	32,69	100
Sociollogy and Political Sciences	14	78,1	21,9	100
Missing**		85,61	14,39	100
Total		62,05	37,95	100

Table 4 - Percentage of professors with at least one paper on ISI by Academic Discipline; ** Missing values on disciplinary area refer mainly to ’91-‘95

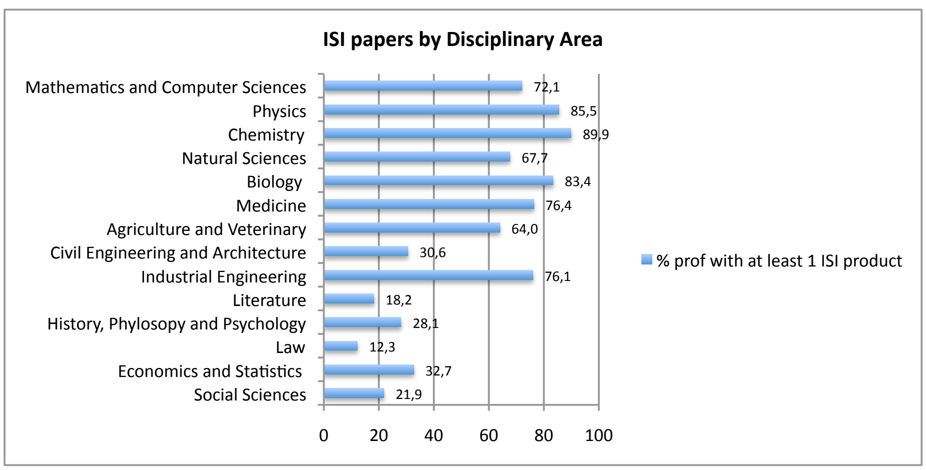


Figure 6: percentage of professors with at least one paper on ISI by academic discipline

4.5 Research Questions and Empirical Strategies

The methodological approach we employ to evaluate the impact of different selection mechanisms caters to the specific research question we ask: "Is there a causal effect of local (vs national) recruitment programs on the subsequent (during the course of next six years) research productivity levels of selected academics?"

We focus on evaluating the effects of a shift to decentralized selection mechanisms in terms of subsequent research productivity six years after average level of the outcome and its time trend. The treatment status can be considered as the exposure of an individual to local selections instead of national ones. The problem is that we can observe almost one of these states for each unit of interest. Indeed individuals who are exposed to local selection programs are by definition (due to a specific time constraint: they were selected after 2000) different from those who are exposed to national recruitment programs. These differences may invalidate the causal comparison of the impact of decentralization on future academics research productivity outcomes.

Recent studies in the econometric literature of program evaluation (Imbens, Wooldridge, 2008) and methodological research on causal inference (Rosenbaum and Rubin, 1983) from observational studies (where investigators have no control over the treatment assignment) suggests the use of propensity scores to accommodate general heterogeneity between two groups of individuals in estimating the treatment effects and to increase precision of the estimates. The treated (selected with local programs) and control (selected with national programs) groups may have significant differences in their observed covariates (scientific discipline and research productivity outcomes) that could lead to biased estimates of the selected effect.

The approach is to estimate the individual probability of an academic being selected in the local program, adjusted for productivity outcome levels in the selection year. Individual research outcomes are essentially measured on three dimensions in the data used in this paper: quantity of publications (given by the cumulative number of papers published in ISI journals), impact of the research (given by the average impact factor of the journals were they published) and the network rate (measured by the total number of citations). Similar levels of estimated probability (measured via propensity scores) signifies similar research productivity propensity for the individual selected with the local program. We can then interpret individuals with equal propensity score as being randomly assigned to national and local programs, in this way reducing the bias due to unbalanced covariates levels between treated and controls.

The study has the characteristics of an experimental design, but lacks randomization over treated and control unit assignment. Such designs are usually classified in literature as quasi-experimental. As underlined by D'Agostino (1998), a question in estimating the propensity score could be: 'Why are we estimating the probability that each individual of the considered groups is selected with a local mechanism (treated) since we already know for certain if it is or not?'.

The answer to this question is that using both nonparametric methods like coarsened exact matching and parametric techniques, such as propensity score matching to adjust our estimate (with any of the most common techniques), we can create a 'quasi-randomized' experiment considering balanced groups of units with respect to their research outcome levels and discipline.

Once estimated, the propensity score at the selection year (that could be considered the starting point) is used to reduce bias in estimating the effect of local recruitment programs over individuals' subsequent research productivity outcomes through widely accepted methodological approaches: regression adjustment (or covariance adjustment), stratification (or sub-classification) and matching. Stratification and regression adjustment techniques are ways to reduce bias in calculating the treatment effect during the estimation procedure (Myers and Louis, 2010), while the matching strategy is used to adjust bias prior to estimation (D'Agostino, 1998).

Our analysis also implements a more robust alternative; coarsened exact matching (CEM) technique (computed via an ad hoc SAS macro language) to our observational study.

Coarsened exact matching is a recent matching method introduced by Iacus, King and Porro (2009) to improve causal inference controlling for the confounding influence of covariates in observational studies. The time-dependent nature of our study, and the flexibility of CEM in estimating non-parametrically two balanced distributions of treated (locally selected professors) and controls (Centrally selected) units, allows us to obtain desired counterfactuals for estimating the decentralization effect.

Balance between the two selected populations is assured by estimating both standard univariate descriptive statistics (as well as most of the commonly presented research papers) and appropriate multidimensional measures of treated and controls distributions.

Similarity in the results of different methodological approaches confirms the robustness of the negative causal effects of decentralization in selection mechanisms for research impact (both final levels and slopes) and international research production (only slopes).

4.5.1 Bibliometric indicators

Three simple bibliometric measures are introduced in this work. Firstly, as quantity of published research we calculate the individual cumulative frequency of ISI items extracted from the database at each year. This measure gives an idea of the quantity of papers on international journals published by the considered researcher up to each considered year.

Secondly, we calculate the cumulative average impact factor of each academic professor at each of the considered years. This measure could be considered as an individual "expected measure of impact".

The second measure is open to criticism. On one hand, one could argue that it is not correct to use journal impact factors to evaluate individual productivity and, on the other hand, that the impact factor could not be considered as a proper measure of quality.

Indeed, impact factors are by definition the arithmetic average of citations of the journal papers in a given period. Referring the average impact factor to each article means that the underlying distribution of citations is assumed to be uniform. But this is empirically unproven, and in some respects refuted by observation. The literature documents that the underlying distribution is strongly asymmetric to the right; few articles receive many citations, and most receive few (Seglen, 1997; Adler, 2009).

It is certainly the case however that the impact factor of a journal provides a priori information about the "expected number of citations" of a published article. In the absence of acceptable citations indexes, we consider the average impact factor of the journals were a researcher has published as a measure of the "expected impact" in terms of citations the researcher will obtain.

Secondly, it is also true that papers published in top journals with high reputation (high impact factor) have already been peer-reviewed by rigorous referees; this process should guarantee high quality standards of the published research. There is thus justification for thinking that a strong correlation exists between journal impact factors and research quality of its published papers.

A measure of network extent could be extracted using the total number of citations. The significance of the network rate measure has to be widely discussed in the literature because of its strong relationship with time. The greater is the number of years an article has been published, the higher is the number of citations due to time exposure on ISI. ISI citations are by definition updated to download time (April, 2012) and they could not be considered comparable measures across years. Researchers selected with national recruitment processes were almost by definition older on ISI in comparison to locally selected individuals. Also controlling citations by the number of years since the first paper on ISI could be misleading because of heterogeneity in ISI behaviours across individuals in the same discipline.

4.6 Empirical Results

An analysis of the descriptive statistics for assistant professors at year of selection (time=-5) evidences the wide differences of bibliometric indicators within academic disciplines. “Hard sciences” professors in Physics or Chemical Sciences at time of their selection had an average of 24 ISI papers, with 2.5 average impact factor, 15 average citations per paper, and more than 400 cumulative citations. On the other end of the spectrum, Arts and Humanities academics have on average less than 1 paper on ISI, with 0.33 impact factor, 1.7 average citations per paper, and 7 cumulative citations in their research careers.

We take log-transformations of research outcomes (impact factor and n° of papers on ISI) to guarantee the normality of both variables distribution. In order to maintain the maximum number of observations, we input zero values for all the individuals with missing values over these variables. The procedure means that when focusing on impact dimension only, we are considering an individual with at least one paper on ISI published on a journal without impact-factor (that is the true reason of its missing on the impact variable). Such a case is coded in the same way as an individual with no paper at all on ISI (missing by absence). This could be a source of bias in our estimates, but we argue the issue is irrelevant for associate and full professors (descriptive below) of hard sciences, although it could be seriously relevant in analyzing assistant professors selection procedures (that are excluded from our analysis).

Id	Disciplinary Area	Freq.	N° ISI papers	Impact factor	Avg. Cit.	Tot. Cit.
1	Mathematics and Comp. Sciences	1282	7,75	1,18	7,07	65,46
2	Physics	1887	24,04	2,62	15,30	402,27
3	Chemistry	1139	24,04	2,99	17,96	500,21
4	Natural Sciences	1644	6,14	1,70	13,75	105,36
5	Biology	1268	16,09	3,43	19,40	344,71
6	Medicine	948	19,47	3,73	14,81	361,47
7	Agriculture and Veterinary	376	6,33	1,38	9,98	90,35
8	Civil Eng. and Archit.	1777	2,21	0,68	4,23	24,64
9	Industrial Engineering	563	10,73	1,43	8,68	107,67
10	Literature	1471	0,70	0,33	1,82	8,89
11	History, Phyl. and Psych.	964	1,49	0,72	3,74	23,24
12	Law	2205	0,43	0,39	1,75	7,41
13	Economics and Statistics	558	1,24	0,66	4,08	14,96
14	Social Sciences	1578	0,62	0,50	1,76	7,79
0	Missing**	279	3,31	0,94	5,50	63,05

Table 5 - Descriptive statistics of bibliometrics indicator by disciplinary area; Associate Professors at year of selection only

Following Kleinman and Horton (2009) and Hedeker and Gibbons (1997), we start by fitting the standard regression model for each of the selected research outcomes (average impact factor, total number of ISI research publications, average number of citations) in the pool of researchers promoted to associate professorship since 1991 up to 2011:

$$ln\left(y_{it}\right) = \beta_0 + \beta_1 time_{it} + \beta_2 treat_i + \beta_3 time_{it} treat_i + \beta_4 areassd_{it} + \beta_5 region_{it} + \varepsilon_{it} \quad (11)$$

where is the intercept that represents the value of the outcome y_i (we use logarithmic transformations of bibliometric indicators to guarantee the normality of the original variable distributions) at the sixth-year, after the pre-reform group of associate professors (controls) has been selected. β_1 is the linear effect of time (research outcome time trend) for the pre-reform group. β_2 is the condition difference between pre- and post-reform associate professors groups at the final year. β_3 is the conditional difference between the two groups, in terms of linear effect of time. β_4 and β_5 are academic discipline and region-specific control variables. Academic disciplines are restricted to hard sciences only, due to the cut-off minimum level of belonging professors with at least one ISI publication.

Parameter	Impact Factor			N° ISI Products			N° Cit.		
	Est.	Std. Err.	Pr>|t|	Est.	Std. Err.	Pr>|t|	Est.	Std.Err.	Pr>|t|
Intercept	0,5549	0,012	<.0001	2,612	0,025	<.0001	2,146	0,021	<.0001
time	0,0037	0,002	0,1246	0,150	0,005	<.0001	-0,003	0,004	0,5395
treat	0,0098	0,009	0,2659	0,383	0,017	<.0001	-0,331	0,015	<.0001
time*treat	-0,0058	0,003	0,0466	-0,043	0,006	<.0001	-0,031	0,005	<.0001
R-squared	0,43			0,24			0,18		
F-Value	1227,2		<.0001	521,35		<.0001	367,03		<.0001
Obs.	44.349			45.611			44.432		

Table 6 - Region and scientific area controls included; log-linear regression model

It would appear that local selection of professors induces better performances in terms of impact (impact factor) and quantity (n° of ISI records) of their research outcomes, and worse productivity differences between two groups in terms of linear effect of time (impact factor only) with respect to the previous national-selected candidates. The time trend appears to be significant only in the second regression model. The average number of citations has a significantly negative treatment and treatment-time interaction estimates associated to post-reform academics. These are not surprising results, as older journal articles have on average a higher number of citations due to longer exposure in comparison to more recent articles.

We use propensity scores since individuals were not randomly assigned to the selection mechanism. Our approach is to select two groups of academics, who, at time of promotion, had balanced observed levels of research outcomes and disciplines. Balancing both treated and control units guarantees less biased comparisons between final (sixth-year) levels and growth rates for the selected outcomes, pre- and post-reform. Propensity score matching is a commonly used technique among economists in observational studies and we start with applying it as our basic model.

With propensity score methods, it is important to check for the overlapping of the propensity score distributions between treated and controls. The univariate reduction imposed by propensity scores avoids the need to check for multidimensional balancing distribution between treated and controls. As discussed in the recent literature (Iacus, 2011), reducing the multivariate distribution into an univariate probability score could be the source of unexpected problems (as the different dimensional spaces of data and propensity score, or the introduction of new bias on few covariates while attempting to balance others).

Treated	Obs.	Mean	Std.Dev	Min	Max
0	5.292	0.6507	0.129	0.0003	0.931
1	12.647	0.7419	0.130	0.0001	0.999

Table 7 - Population propensity scores descriptive statistics

Table 7 reports that mean propensity is larger in the post-reform group in comparison to the pre-reform group, meaning that the fitted standard logistic model is a quite good predictor of the treatment status, and generates an effective estimate of individuals propensity score. However, maximum and minimum values of the propensity are similar for the two groups. To check for the goodness of fit for the overlap, we propose (Figure 7) histograms of propensity score distributions for each of the two groups.

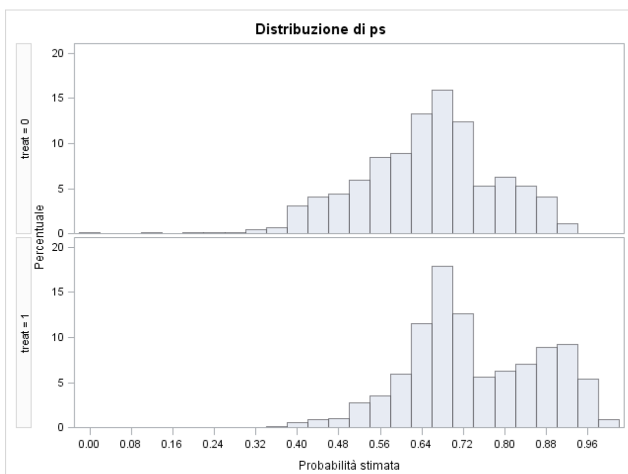


Figure 7: Propensisty Scores histograms by treatment and control groups

To be sure that there is perfect overlapping of the propensity scores distributions between the two groups (and that we are not extrapolating outside the range of data when we adjust the model for propensity), we cut the distribution (both to the left and to the right) and we exclude all the units with propensity lower than 0.42 and higher than 0.94, so that a common support is identified. The number of individuals considered decreases of about 600 units, and the resulting sample is composed of 5.071 assistant professors selected before the reform, in comparison to 12.371 individuals selected after reform, with more balanced propensity distributions.

Treated	Obs.	Mean	Std.Dev	Min	Max
0	5.071	0.691	0.051	0.42	0.94
1	12.371	0.717	0.073	0.42	0.94

Table 8 - Common support propensity scores descriptive statistics by treatment group

If we fit treatment effect on the restricted balanced distributions, a slight difference appears in the estimates of treated academics with respect to controls (increasing on the impact factor and decreasing on the number of ISI records), with a still significant evidence of impact factor and n° of ISI papers treatment associations on the interaction with time. Goodness of fit is also slightly improved.

Parameter	Impact Factor			N° ISI Products			N° Cit.		
	Est.	Std.Err.	Pr>|t|	Est.	Std.Err.	Pr>|t|	Est.	Std.Err.	Pr>|t|
Intercept	0,581	0,012	<.0001	2,791	0,024	<.0001	2,180	0,022	<.0001
time	0,008	0,002	0,0006	0,186	0,005	<.0001	0,007	0,004	0,082
treat	-0,001	0,009	0,8971	0,268	0,017	<.0001	-0,343	0,015	<.0001
time*treat	-0,007	0,003	0,0099	-0,063	0,005	<.0001	-0,033	0,005	<.0001
R-squared	0,45			0,26			0,19		
F-Value	1227,2		<.0001	546,55		<.0001	373,49		<.0001
Obs.	43.313			43.313			42.600		

Table 9 - Region and scientific area controls included; common support log-linear regression model

We estimate the treatment effect with three techniques using the propensity score as a means of reducing differences due to potentially confounding covariates in the selected individuals.

4.6.1 Propensity Score Regression Adjustment

In clinical research (where observational studies are widely diffused) the regression adjustment technique is one of the most commonly used methods (Shah et al., 2005; Weitzen et al., 2005, Stuart et al., 2010).

Table 10 shows the results of fitting this first technique on the restricted distributions, including propensity scores as covariates in ***a-la Heckman*** way (1979) to account for individual heterogeneity. Results underline significant negative associations between local selection mechanisms (treated) and both impact factor and number of ISI products at the sixth-year after selection. Region and scientific discipline are included as geographic and 'homogeneity' controls. The time trend is significant with a positive effect for all the models fitted, and interactions with time are both significant and negative for impact and quantity of papers. A conditional difference between the two groups of academics in terms of linear effect of time at the 10% significance level has a negative association, considering the impact and ISI quantity measures.

Parameter	Impact Factor			N° ISI Products			N° Cit.		
	Est.	Std.Err.	Pr>|t|	Est.	Std.Err.	Pr>|t|	Est.	Std.Err.	Pr>|t|
Intercept	0,420	0,014	<.0001	1,635	0,025	<.0001	2,234	0,025	<.0001
time	0,008	0,012	0,0003	0,187	0,004	<.0001	0,007	0,004	0,0856
treat	-0,054	0,009	<.0001	-0,112	0,016	<.0001	-0,034	0,016	<.0001
time*treat	-0,006	0,002	0,0176	-0,058	0,005	<.0001	-0,034	0,005	<.0001
ps	0,275	0,012	<.0001	1,989	0,022	<.0001	-0,091	0,022	<.0001
R-squared	0,45			0,37			0,20		
F-Value	1.240		<.0001	900		<.0001	361		<.0001
Obs.	42.483			43.313			42.600		

Table 10 - Region and scientific area controls included; common support log-linear regression model

We also fit log-linear regressions by academic discipline, stressing as the main result the presence of substantial heterogeneity between them, with most of the hard-sciences reflecting the negative effects with few exceptions.

Log-Linear Regression model					
Area	**Parameter**	**Impact Factor**		**N° ISI Papers**	
	Variables	**Est.**	**Pr>|t|**	**Est.**	**Pr>|t|**
1	treat	-0,0077	0,792	0,0335	0,4131
	time*treat	-0,0071	0,446	-0,0555	<.0001
2	treat	-0,0546	0,112	-0,1882	0,004
	time*treat	-0,0034	0,743	-0,0743	0,0002
3	treat	0,0641	0,0005	-0,1938	<.0001
	time*treat	0,0007	0,9074	-0,0687	<.0001
4	treat	-0,1940	<.0001	-0,0584	0,5073
	time*treat	-0,0260	0,074	-0,0283	0,314
5	treat	0,0212	0,3769	-0,2358	<.0001
	time*treat	-0,0054	0,4659	-0,0669	<.0001
6	treat	-0,2123	<.0001	-0,2988	<.0001
	time*treat	-0,0188	0,0117	-0,0740	<.0001
7	treat	-0,0826	0,0118	-0,0443	0,3984
	time*treat	-0,0118	0,2574	-0,0313	0,0599
9	treat	-0,0180	0,3196	0,0837	0,0181
	time*treat	0,0020	0,7235	-0,0465	<.0001

Table 11 - Region control included; common support log-linear regression model, hard sciences only

4.6.2 Quantile treatment regression

An alternative method of estimating treatment effects implements the same strategy, but stratifies by propensity score quartile. Stratification has only recently been applied in the economics literature (Bitler, Gelbach, and Hoynes, 2002; Abadie, Angrist and Imbens, 2002), although in statistics the approach has been studied since the early seventies. Following this methodology, propensity scores are split into different strata (usually following distributional quantiles) and the treatment effect is estimated within each stratum. The general effect on treated units is then calculated as the weighted mean of the stratum-specific obtained estimates.

Quartile	Parameter	Est.	Std.Err.	tValue	Pr>|t|	R-squared	Pr>F
1st	Intercept	0,474	0,023	20,02	<.0001	0,41	<.0001
	time	0,014	0,004	3,66	0,003		
	treat	-0,008	0,003	-2,11	0,034		
	time*treat	-0,006	0,005	-1,06	0,289		
2nd	Intercept	0,548	0,035	15,56	<.0001	0,45	<.0001
	time	0,006	0,007	0,88	0,381		
	treat	-0,050	0,024	-2,11	0,034		
	time*treat	-0,003	0,007	-0,41	0,681		
3rd	Intercept	0,610	0,019	31,89	<.0001	0,49	<.0001
	time	0,003	0,003	0,89	0,374		
	treat	-0,027	0,013	-1,94	0,052		
	time*treat	-0,002	0,005	-0,36	0,721		
4th	Intercept	1,566	0,129	12,12	<.0001	0,38	<.0001
	time	-0,023	0,041	-0,57	0,569		
	treat	-0,0823	0,127	-6,47	<.0001		
	time*treat	0,022	0,041	0,54	0,590		

Tab 12 - Region and scientific area controls included; common support; log-linear model of impact factor

The upper and following tables (12 and 13) stress the extent to which the treatment effect is significant in all four quartiles with a negative estimate for the impact model, while only the first quartile reported a negative overall estimate statistically significant for n° of ISI papers.

Quartile	Parameter	Est.	Std.Err.	tValue	Pr>|t|	R-squared	Pr>F
1st	Intercept	2,319	0,042	55,78	<.0001	0,28	<.0001
	time	0,226	0,006	34,94	<.0001		
	treat	-0,175	0,031	-5,64	<.0001		
	time*treat	-0,005	0,010	-0,53	0,5946		
2nd	Intercept	3,017	0,062	48,30	<.0001	0,22	<.0001
	time	0,190	0,012	15,67	<.0001		
	treat	-0,043	0,042	1,01	0,3122		
	time*treat	-0,021	0,014	-1,53	0,1256		
3rd	Intercept	3,150	0,036	86,63	<.0001	0,24	<.0001
	time	0,120	0,007	17,18	<.0001		
	treat	-0,009	0,026	-0,36	0,7219		
	time*treat	-0,006	0,008	-0,67	0,5017		
4th	Intercept	3,680	0,215	17,14	<.0001	0,20	<.0001
	time	-0,084	0,069	1,21	0,228		
	treat	-0,334	0,212	-1,58	0,114		
	time*treat	0,014	0,069	-0,21	0,834		

Tab 13 - Region and scientific area controls included; common support; log-linear model of number of paper ISI

The results also show that the time trend is statistically significant with a positive effect in the first quartile of the impact distribution, while treatment effect is significant to the overall distribution with negative coefficients that means lower average values at the sixth-year after selection.

In terms of ISI products the picture changes, a positive time trend with different effects persists along all quartiles, and a negative treatment effect is not statistically significant except for the first bottom 25% of associate professors. No slope is significantly different between post- and pre-reform researchers. It appears to be the case that local selection procedures are associated with lower levels of ISI products rates for the bottom 25% of academics.

It would also seem that local recruitment procedures induce negative incentives on lower quality researchers' production and on sixth-year quantity/quality levels of all the academic population. Good and top researchers are invariant in their incentives in producing a number of high quality papers on international journals.

Finally, we note that these results could be influenced by the choice of the number of propensity score strata made we use. This choice usually influences the bias of the resulting estimates. Generally the effect is twofold, as the width of the strata produces effects on both variance and bias of the estimates. Wide strata are associated with lower levels of variance but higher bias, while narrow strata produce high variance but lower bias on the estimates. Otherwise, Cochran's (1968) empirical results provide evidence of that 90% of the initial bias is eliminated by quantile stratification.

4.6.3 Propensity Score Matching

We then match between treated and controls units three times according to the propensity score distance choice: of 0.001, 0.005 and 0.0001. We write an SAS macro to combine individuals with propensities in a pre-determined interval limit. When one individual is assigned as control of one treated, both are automatically excluded from the possible-matching individuals by the macro. This is a selection without replacement technique, based on propensity differences. The aim is to obtain two subsamples of treated and control units with the same probability of being treated, according to individual bibliometric levels and academic discipline at the individual's career starting point (fixed as the year of selection for all individuals).

Three pairs of groups are selected in this manner. Results from the different sensitivity levels are reported in the table 14 to assure that matching invariance is present in our results, and to indirectly check for estimate robustness. We know from literature and recent studies that good matching can be highly effective in removing imbalance in covariates between treatment and control groups, allowing researchers to reduce the bias due to individual differences and estimate unbiased (or less biased) treatment effects over the identified matching units.

Parameter	treat	log impact factor			log n° ISI products		
		Matching			Matching		
		0.001	**0.005**	**0.0001**	**0.001**	**0.005**	**0.0001**
time		0,0058***	0,0028**	0,0031**	0,168***	0,1697***	0,1585***
		[0,0013]	[0,0013]	[0,0014]	[0,0021]	[0,002]	[0,0023]
treat	0	0,6426***	0,6791***	0,7107***	3,484***	3,4621***	3,496***
		[0,0746]	[0,0758]	[0,0821]	[0,0461]	[0,0461]	[0,0522]
treat	1	0,6589***	0,6928***	0,7265***	3,537***	3,5744***	3,5716***
		[0,0745]	[0,0018]	[0,0822]	[0,0444]	[0,044]	[0,0505]
time(treat)	0	0,0064***	0,0090***	0,0086***	0,0051 *	0,0056**	0,0061**
		[0,0019]	[0,0018]	[0,0021]	[0,0029]	[0,0029]	[0,0032]
time(treat)	1	.	.	.	.	.	.
treat a vs. b intercept		0,0163	0,0137	0,0157	0,0529 *	0,1123***	0,0755**
		[0,0167]	[0,0158]	[0,0184]	[0,0316]	[0,030]	[0,0345]
treat a vs b slope		-0,0064***	-0,0090***	-0,0086***	-0,0051*	-0,0056**	-0,0061**
		[0,0019]	[0,0018]	[0,0021]	[0,0029]	[0,0029]	[0,0032]
Pr > ChiQuadr		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
Obs.		27.708	53.361	22.190	29.817	31.403	24.144

Tab 14 * significant at 10%; ** significant at 5%; *** significant at 1%, standard errors in brackets, region and scientific discipline controls included

Results show a negative and significant effect in treatment slopes both for impact of research and quantity. An average positive treatment effect is evident for the quantity of paper published by associate professors at six-year after selection.

Absolute values of treatment and slope estimates do not change, allowing for more precise levels of matching between treatment and control units. This means that our matching methods have acceptable performance, with less variance in the selected units. Figures below (8 and 9) provide evidence of slope patterns for impact and quantity of research in each of the matching interval.

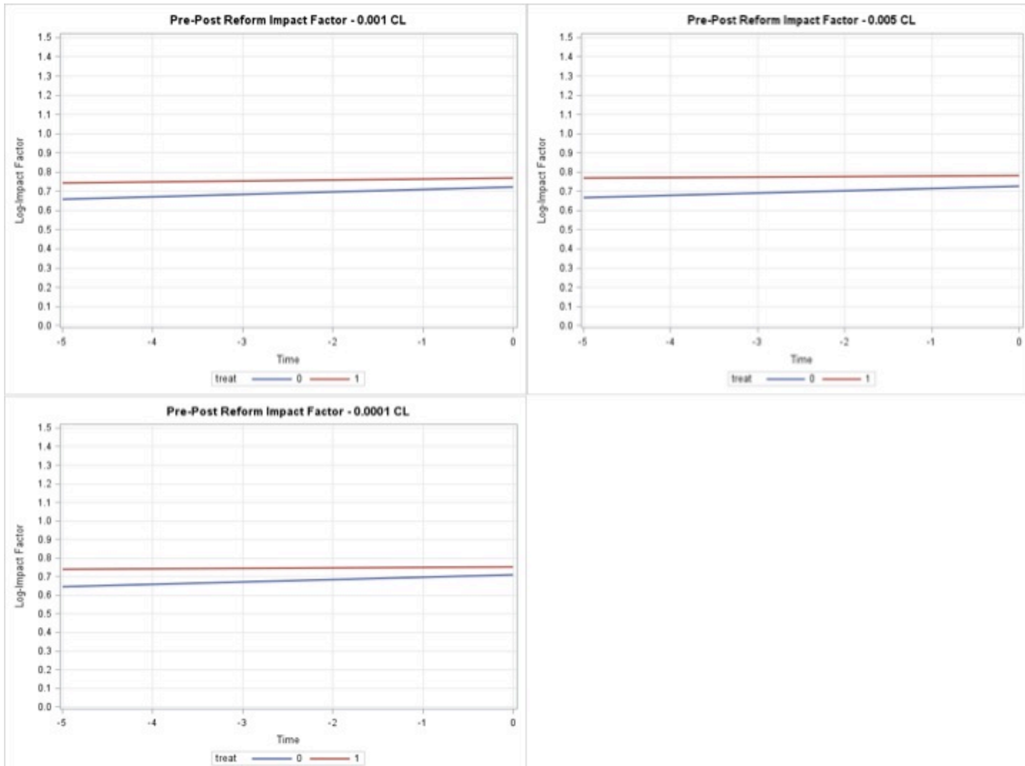


Figure 8: Growth patterns of Associates research impact at different matching thresholds

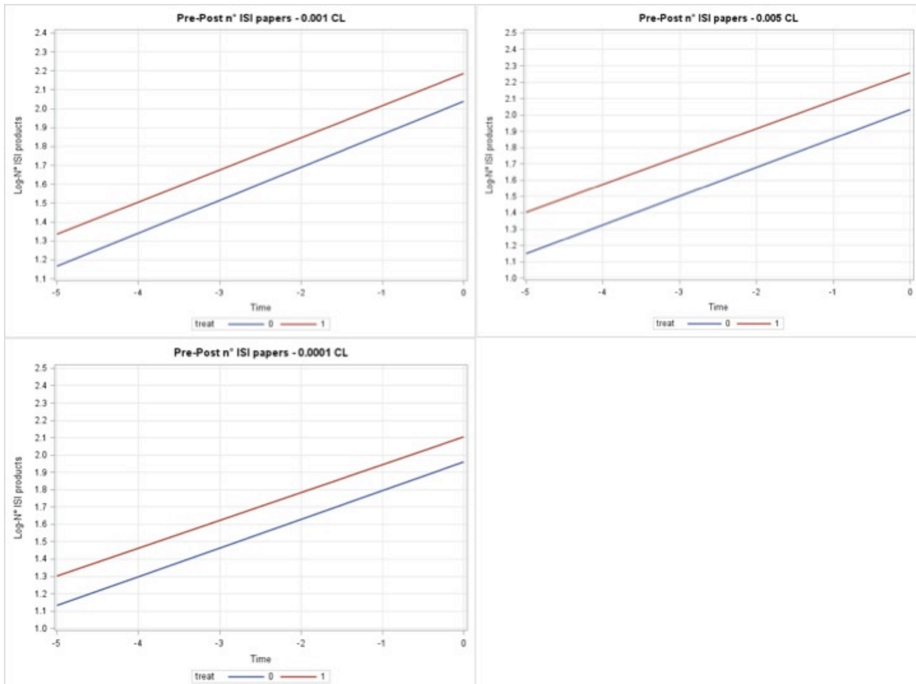


Figure 9: Growth patterns of Associates research productivity at different matching thresholds

4.6.4 Coarsened Exact Matching

The main goal of CEM (Iacus S. et al., 2007 and 2009) is to balance the empirical distributions between treated and control groups, obtaining exactly balanced data. Balanced groups avoids having the researcher control for the heterogeneity while specifying the model, meaning that simple differences in means are good estimates of the causal effect. But usually finding a matching solution in empirical propensity applications does not guarantee good balance to all the selected covariates. Improving balance on most of them could leave the remainders imbalanced, often introducing more bias with respect to the initial distribution.

In addition to this, propensity score matching (and Mahalanobis distance methods) has the drawback of violating the congruence principle, which requires congruencies between data and analysis spaces metrics (the own metric of the two spaces is different). Parametric methods usually force covariates of the original data from a multi-dimensional original space in a new space defined by the propensity itself. Mielke and Berry (2007) show how violating this principle produces less robust inferences.

In comparison, coarsened exact matching meets the principle of not reducing the original data space, operating in the multidimensional variable space itself. The novelty in CEM in comparison to other matching methods is the approach's reverse structure: setting the balance between treated and controls first and then get the matched units as result (sample size is not a priori known). Common previous techniques proceed as reverse guaranteeing final matched sample size ex ante, and usually then reducing imbalances between treated and controls.

The structure of this recent method consists of sets of strata for each variable, sorting them into groups and then applying exact matching techniques to each stratum of treated and controls finding pairs of matched units. Strata without at least one treated and one control units are pruned from the sample. Matching could be one-to-one, simply randomly selecting the same number of treated and controls in each stratum. Otherwise, in both cases of no possible matches and multiple matches of treated and controls between the same strata, CEM assigns a weight to each unit. The weights are 0 to unmatched, 1 to treated and the relative weight of the treated and controls of each stratum with respect to the corresponding weight in the entire population to the matched units. The remaining differences in balance (due to the coarsening choice at strata level) are then spanned by introducing statistical models (weighted regressions, within strata models. . .) without serious risks of great model dependence.

The advantage of CEM is that it allows for removing, with a single step procedure, all covariate imbalances between controls and treated matching units by coarsened strata. The only possible distributional difference after CEM is caused by the a-priori chosen level of coarsening. Indeed, CEM is a data-analyst dependent procedure that follows the own researcher choices on variables coarsening. CEM is hence a useful nonparametric tool in estimating the counterfactual potential outcome of each unit for making causal inference in observational studies.

Applying CEM to our study means firstly to set variable-by-variable the non-overlapping intervals to coarsen original data about winners of selection procedures (before and after the reform) at the year of their selection. This is to match one-to-one each stratum treated and controls units after removing all the individuals (treated and controls) owning to zero controls strata. We write a SAS macro procedure that select only full and associate at the year of their selection, creating strata for each combination of the coarsened covariates. Available covariates regarding bibliometric indicators and disciplinary area are coarsened according to reasonable assumptions. A 0.3 impact factor intervals, a one-to-one n° of ISI publications and a 10 pages interval of cumulative number of pages written by the authors are settled as coarsening rules. Missing data are treated as 'missing as zeros' due to the particular nature of our data. Missing values of bibliometric indicators (impact factor, n° of paper ISI, citations, sum of pages etc...) reflect the absence of the author in the data and absence on ISI is equal to 0 international papers published, with 0 pages written, 0 citations received and 0 average impact factor. A real drawback of missing replaced with zeros could be represented by the equal 'treatment' of an author with few ISI publications with zero impact factor and 0 citations and an author without ISI records. However, having zero ISI publications or few records with no impact factor and no citations at the associate professorship level in our restricted word (hard sciences only) could be considered, without a significant loss of information, to be quite the same. So replacing missing values of bibliometric indicators with zeros seems to be useful to our purposes. The desirable output of this procedure is a sample of balanced treated and controls. For this case, we found 3.181 treated professors with one-to-one coarsened exactly matched controls over 5.292 potentially possible 1:1 couples.

Groups		Frequencies	Sample Frequencies
Treated	total	12.646	3.181
	zero controls strata	6.096	
Untreated	total	5.292	3.181
	zero controls strata	1.727	
		matched units	6.362

Table 15 - Frequencies of Treated and Untreated units by CEM groups

The selected sample population is now composed of comparable sub-groups of individuals (selected before and after the reform) with similar levels of bibliometric indicators (according to the coarsened intervals settled as before) and operating their research effort in the same disciplinary areas. We estimate the average reform effect on both the sixth-year research productivity and difference in slopes between before and after reform associate professors.

The following table (Tab.16) shows statistics for three of the four selected variables, impact factor, number of paper ISI and cumulative sum of written pages. Frequencies and descriptive statistics reported underline the differences between the two sub-populations (most of the difference is plausibly due to the different time horizon at which the two populations refer to).

treat=0					
Variable	**N**	**Min**	**Mean**	**Max**	**StdDev**
Impact factor	5.292	0	1,28	25,285	1,954
N° of paper Isi	5.292	0	5,00	133	10,156
Sum of written pages	5.292	0	35,72	923	67,93
treat=1					
Variable	**N**	**Min**	**Mean**	**Max**	**StdDev**
Impact factor	12.647	0	1,65	53,48	2,245
N° of paper Isi	12.647	0	9,55	231	16,334
Sum of written pages	12.647	0	253,1	923	5581

Table 16 - Descriptive statistics by treated and controls

Table 17 provides evidence (univariate absolute difference in means) of balance between CEM selected treatment and control groups in the overall sample. Mean and standard deviations of the two, equal-size, samples of units are relatively close from one to the next. Before and after reform associate professors have an average number of 1.2 papers published on ISI journals with an average impact factor of 0.45 and a n° of pages close to 10.

treat=0					
Variable	**N**	**Min**	**Mean**	**Max**	**StdDev**
Impact factor	3.181	0	0,459	10,61	0,998
N° of paper Isi	3.181	0	1,201	41	3,247
Sum of written pages	3.181	0	9,492	249	24,793
treat=1					
Variable	**N**	**Min**	**Mean**	**Max**	**StdDev**
Impact factor	3.181	0	0,457	10,55	0,995
N° of paper Isi	3.181	0	1,222	49	3,365
Sum of written pages	3.181	0	9,801	255	27,496

Table 17 - Descriptive statistics of matched units by treated and controls

By construction, covariate descriptive statistics over the entire sample are almost equal to descriptive statistics of units in each of the selected disciplinary areas. However, despite being commonly use in observational studies (especially in propensity score studies), univariate distributions of means do not guarantee the absence of bias in estimating the treatment effect. Recent studies (Iacus et al, 2011) looking at the multidimensional histograms of the two samples (for treated and controls) introduce methods to check for multivariate balancing of their empirical distributions. They propose as measure of imbalance the L_1[25] that is the semi-sum of the absolute differences between relative frequencies of treated and controls for each identified strata.

$$L_1(f, g) = \frac{1}{2} \sum_{l_1 \ldots l_k \epsilon H(X)} |f_{l_1 \ldots l_k} - g_{l_1 .. l_k}| \tag{12}$$

L_1 for the entire population is close to 1 (highly unbalanced distribution of treated and controls). This means that a substantial number of cells in the multidimensional matrix have zero controls (or treated). Comparing the L_1 of the matched population with the previous one provides evidence of the unbalanced reduction due to CEM. L_1 is equal to 0.19 after CEM, this means high rate of balancing between the populations of treated and controls (Table 18).

L1 - matched	L1 - population
0.19	0.92

Table 18 - L1 matched and original population multivariate balance measures

We then plot parallel coordinates plot as a visualization method for detecting patterns of matched and unmatched units in a multivariate setting. Looking at graphs 10 and 11[26] it appears that the full professor matching individuals are relatively well distributed between the considered dimensions; they belong to all the academic disciplines, produce a number of papers, with average citations and impact factor on ISI in the first bottom half of the distribution. On the

[25] as defined by Iacus et al. (2011) if $H(X_1)$ is defined as the set of distinct values generated by binning on variable X_1 then the consequent multidimensional histogram is constructed from the set of cells generated by the Cartesian product $H(X_1) \times ... \times H(X_k)$. Let f and g be defined also as the relative empirical frequency distributions for the treated and control groups then $L_1(f, g)$ is defined by the proposed equation. The L_1 takes values between 0 and 1: if the treated and controls distributions are completely separated then $L_1 = 1$, if they perfectly overlap then $L_1 = 0$. In middle cases $L_1 \epsilon (0, 1)$.

[26] Thanks to the open source GGobi software

other side of the story matched assistant professors are distributed as well in the first bottom half of the distribution between these dimensions.

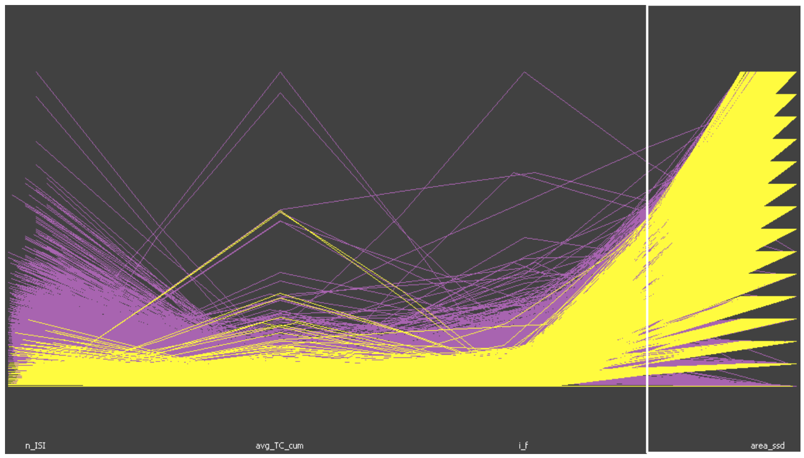


Figure 10: Full Professors parallel coordinates plot across included covariates

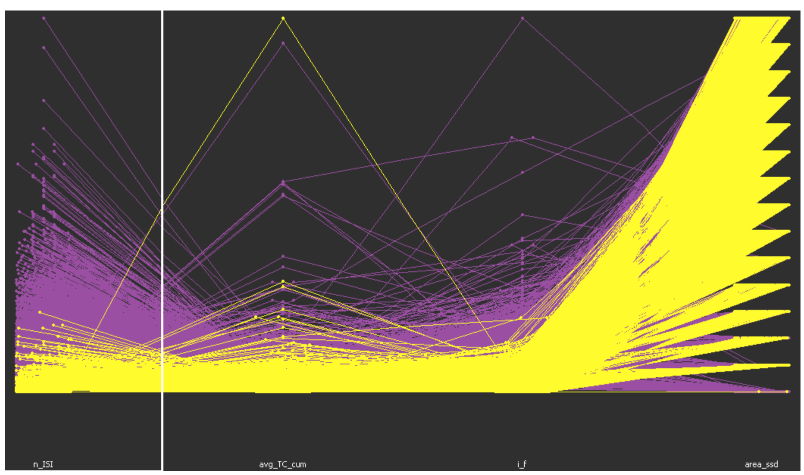


Figure 11: Associates parallel coordinates plot across included covariates

We then fit the a growth longitudinal model (Hedeker and Gibbons 1997; Overall et al. 1999; Howell et al., 2008)) through the individual careers of treated and controls, that allow us to identify a negative effect of decentralization on research slopes (impact and quantity) of selected individuals, for both associate and full professors. Our motivation for using also a growth model (also called unbalanced repeated measures model) is firstly because it is specifically designed for exploring longitudinal structure of the data. Moreover, in commonly specified generalized linear models, estimations are carried out forcing the pattern of covariances (or correlations) to be constant across time. This means that we require that all subjects in each group to change in the same way over time. This is an unrealistic assumption especially thinking of the heterogeneity of research patterns across individuals. "The second reason to propose also a growht model is that it allows for different lengths of measurements for different subjects. The model estimates the subject's trend across time on the basis of whatever data that subject has, augmented by the time trend that is estimated for the sample" (Hedecker and Gibbons, 1997). Thus inference is based on all available measures included in the data. Time can also be continuous, rather than a fixed set of points and is modeled as random. First, let us consider a model of the changes in logarithmic transformation of research output yi across time as a function of treatment group, time and the interaction of treatment and time:

$$ln\left(y_{it}\right) = \beta_0 + \beta_1 time_{it} + \beta_2 treat_i + \beta_3 time_{it} treat_i + v_{0i} + v_{1i} time_{it} + \varepsilon_{it} \quad (13)$$

where $ln\left(y_{it}\right)$ is the log-transformation of research output (impact factor, n° of paper on ISI etc). β_0 is the mean of the dependent variable of all the selected individuals at time zero (sixth year). The term $time_{it}$ is a time variable with values from -5 (at time of selection) to 0 (at the sixth year after selection) for each i-th individual.

So, in terms of its representation, this model could be divided into a within-subject model:

$$ln\left(y_{it}\right) = b_{0i} + b_{1i} time_{it} + \varepsilon_{it} \quad (14)$$

and the between-subjects model:

$$b_{0i} = \beta_0 + \beta_2 treat_i + v_{0i} \quad (15)$$

$$b_{1i} = \beta_1 + \beta_3 treat_i + v_{1i} \quad (16)$$

With this econometric characterization of treatment and time effects, we can interpret the parameters as follows: β_0 is the average of $ln\,(y_i)$ at time 0 (sixth year after selection) for the untreated group (nationally selected individuals, where treat=0), β_1 is the average trend across time for the untreated group (treat=0); β_2 is the average difference in $ln\,(y_i)$ at the sixth-year after selection between centrally and nationally selected professors; β_3 represents the average difference in trend lines between treated and untreated. Furthermore this regression model allows each individual to deviate from the owning group trend line in terms of final intercept (v_{0i}) and time-trend across time (v_{1i}).

Then the first step is to estimate the Unconditional Means Model (UMM) and the Unconditional Growth Model (UGM) to quantify respectively the output variability between individuals (without taking into account the effect of time) with the former (UMM) and also within the individuals over time with the latter (UGM). These simple models allow to establish if a systematic output variation is present in our data and if it is between and/or within subjects.

The UMM is specified as follows:

$$ln\,(y_{it}) = \pi_{0i} + \varepsilon_{it} \tag{17}$$

$$\pi_{0i} = \gamma_{00} + \xi_{0i} \tag{18}$$

where $\varepsilon_{it} \sim N(0, \sigma_\varepsilon^2)$ and $\xi_i \sim N(0, \sigma_0^2)$, π_{0i} is the mean output (impact factor in this case) of subject i (*specific mean*), γ_{00} is the mean output of all the subjects involved, so it is the mean of the individuals' output means (*grand mean*). Table 19 shows results for UMM model of Associates Professors. Random effects estimates of covariance parameters are significantly different from zero and the output mean varies both within and between subjects. The intraclass correlation coefficient, ρ , give us the individuals variance component.

$$\rho = \frac{\hat{\sigma_0^2}}{\hat{\sigma_0^2} + \hat{\sigma_\epsilon^2}} = \frac{0.96}{0.96 + 0.40} = 0.68 \tag{19}$$

The 68% of the total output variation is due to differences between individuals consequently the 32% is due to the effect of time and individual characteristics. The intraclass correlation coefficient is also a measure of the residual autocorrelation of the model, so that in our case the individual average autocorrelation of each couple of residuals is 0.68 (really far from the absence of autocorrelation in OLS).

Covariance Parameters Estimates - UMM					
Cov. Parm	Subject	Estimate	Std. Error	Z Value	Pr. Z
AR(1) - $\hat{\sigma_0^2}$	Id	0.961	0.0014	679.6	<.0001
Residual - $\hat{\sigma_\varepsilon^2}$		0.408	0.0133	30.6	<.0001
Fixed Effects Solution					
Effect	Estimate	Std. Error	DF	t Value	Pr > |t|
Intercept - $\hat{\gamma_{00}}$	0.550	0.0151	1621	36.3	<.0001

Table 19 - Unconditional Means Model estimates for log of impact factor; selected associates professors

When estimating the UGM model (including time as unique regressor) the $Pseudo - R^2 = \frac{\sigma_\epsilon^2(UMM) - \sigma_\epsilon^2(UGM)}{\sigma_\epsilon^2(UGM)}$, that give us a measure of the variance proportion explained by the inclusion of the time regressor, is 0.96. It means that the 96% of the within subject variability is due to the effect of time. Consistent estimates of the UMM and UGM models derive using the other output measures and for full professors.

Looking now at the results of the completed Growth Model significant negative effects regarding local selections on both final outcomes level (β_2) and slope differences (β_3) are estimated for impact factor outcomes in both career steps. Local selection negative effects are statistically significant only for slope differences with the quantity measure of international research. Associates Professor estimations of treatment effects are shown in table 20. Robustness checks are shown for two different cases: a "pure incentives" model (Pre '94 after '00) where all the selected individuals were sure of the selection rules (non changes were predictable during the previous years) and a "restricted" model (restricted to the five years before and the five years after the reform) were the hypothesis of more similar quality distributions of treated and controls researchers is largely acceptable.

		Growth model		Pre '94 After '00 growth model		'96-'05 Growth model	
Parameter	treat	impact factor	n° ISI	impact factor	n° ISI	impact factor	n° ISI
Intercept		0,302***	2,197***	0,552***	1,98***	0,267***	2,27***
		[0,035]	[0,066]	[0,044]	[0,072]	[0,046]	[0,082]
time		-0,026**	0,126***	-0,011	0,21***	-0,011	0,16***
		[0,010]	[0,020]	[0,016]	[0,026]	[0,012]	[0,025]
treat	0	0,077***	-0,051	0,137***	-0,201***	0,042***	-0,159
		[0,017]	[0,034]	[0,027]	[0,043]	[0,020]	[0,043]
treat	1	.	.	.	.	.	.
time(treat)	0	0,081**	0,019***	0,008	0,035	0,046***	0,115***
		[0,011]	[0,020]	[0,027]	[0,043]	[0,014]	[0,028]
time(treat)	1	.	.	.	.	.	.
treat a vs b intercept		-0,077***	0,051	-0,137***	0,201***	-0,042***	0,159***
		[0,017]	[0,034]	[0,027]	[0,043]	[0,020]	[0,043]
treat a vs b slope		-0,081**	-0,019***	-0,008	-0,035	-0,046***	-0,115***
		[0,011]	[0,020]	[0,027]	[0,043]	[0,014]	[0,028]
Pr > ChiQuadr		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
Obs.		6.362	6.362	5.111	5.111	5.151	5.609

Table 20 - *significant at 10%; ** significant at 5%; *** significant at 1%, standard errors in brackets, Associate Professors model with region and scientific discipline controls included

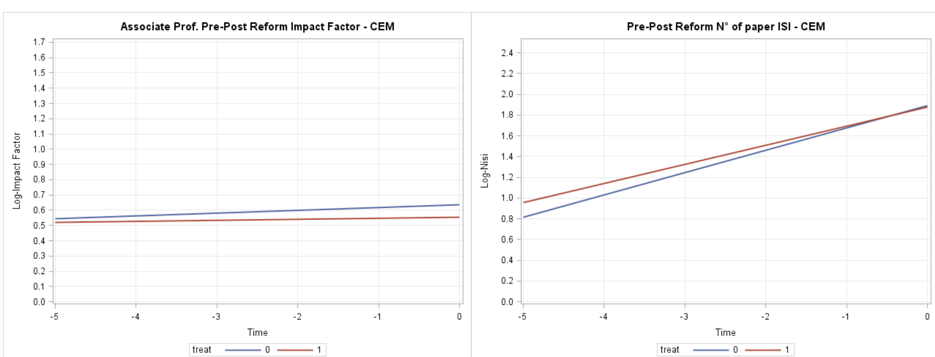


Figure 12: CEM - Associates' impact and quantity growth lines by treatment and control groups

Table 21 provides evidence of a treatment effect on the careers of full professors (sixth year after selection). Both final research impact and impact slopes are statistically significant and negative.

		Growth model	
Parameter	**treat**	**log impact factor**	**log n° ISI products**
time		0,014***	0,159***
		[0,044]	[0,0066]
treat	0	0,129***	0,014
		[0,0465]	[0,087]
treat	1	.	.
time(treat)	0	0,014**	0,051***
		[0,0063]	[0,0094]
time(treat)	1	.	.
treat a vs b intercept		-0,129***	-0,014
		[0,0465]	[0,087]
treat a vs b slope		-0,014**	-0,051***
		[0,0063]	[0,0094]
Pr > ChiQuadr		<.0001	<.0001
Obs.		3.090	2.763

Table 21 - * *significant at 10%; ** significant at 5%; *** significant at 1%, standard errors in brackets, Full Professors model with region and scientific discipline controls included

Newly hired full professors have fewer incentives to produce high quality papers in the subsequent years. The "number of papers" outcome highlights a negative difference in slopes (better for untreated) and no difference in the final quantity level.

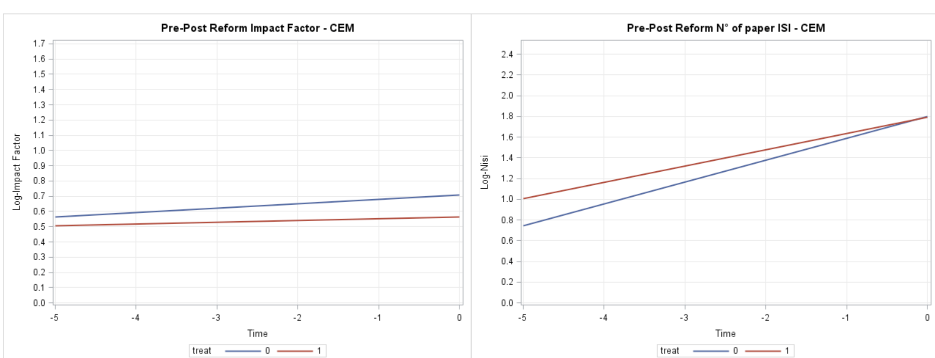


Figure 13: CEM - Full Professors' impact and quantity growth lines by treatment and control groups

Finally, one could also argue that estimating the average effect of decentralized selection mechanism (our treatment status) over a subset of exactly matched associate professors (that theoretically is equivalent to restricting to the common support in propensity score matching) does not consider all the available information.

Considering the fact that we think that if no matching is available is better to prune these units instead of artificially include them with one of the possible statistical tricks. Our objection is the following: extrapolating a model fitted only on matched units to predict counterfactuals for the unmatched is still twisting because pruned units are predicted with a model estimated on the common support only (we could not be sure of betas' validity above the overall population). However we show, using a standard extrapolation technique, that our results are still robust, even allowing for the inclusion of unmatched units. We extrapolate the model estimated on matched units, and fit these over unmatched units, obtaining counterfactuals. The average treatment effect is the weighted average of matched and unmatched estimations.

The model results by each of the selected disciplines shows the negative effects of local recruitment process over Math and Computer Science, Earth Sciences, Medicine and Agricultural and Veterinary in terms of final impact factor level (sixth year after selection) of associate professors. The slope is negative and significant only in the Match and Computer Science area. Both estimates of the standard regression model and growth model are basically consistent (Table 22).

Academic Discipline		Regression Model			Growth Model		
		Est.	Std.Err.	Obs.	Est.	Std.Err.	Obs.
Mathematics and Computer Sciences	treat	-0,13***	0,042	537	-0,15***	0,034	537
	time*treat	-0,013	0,014		-0,014***	0,015	
Physics	treat	0,22	0,133	151	0,31	0,17	151
	time*treat	-0,057	0,042		-0,036	0,12	
Chemistry	treat	0,03	0,035	348	0,021	0,049	348
	time*treat	-0,0039	0,011		-0,024	0,042	
Natural Sciences	treat	-0,22**	0,099	270	-0,17***	0,11	270
	time*treat	-0,049	0,035		-0,038	0,09	
Biology	treat	0,04	0,014	620	0,085	0,071	620
	time*treat	-0,015	0,061		-0,024	0,068	
Medicine	treat	-0,11***	0,044	1703	-0,12****	0,045	1703
	time*treat	-0,015	0,015		0,025	0,043	
Agriculture and Veterinary	treat	-0,19**	0,013	1371	-0,20**	0,051	1371
	time*treat	-0,037***	0,051		0,0143	0,029	
Industrial and Information Engineering	treat	-0,028	0,029	2728	-0,011	0,028	2728
	time*treat	0,003	0,011		0,007		

Table 22 - * * significant at 10%; ** significant at 5%; *** significant at 1%, standard errors in brackets, Log of impact factor Associate Professors model by discipline with region controls included.

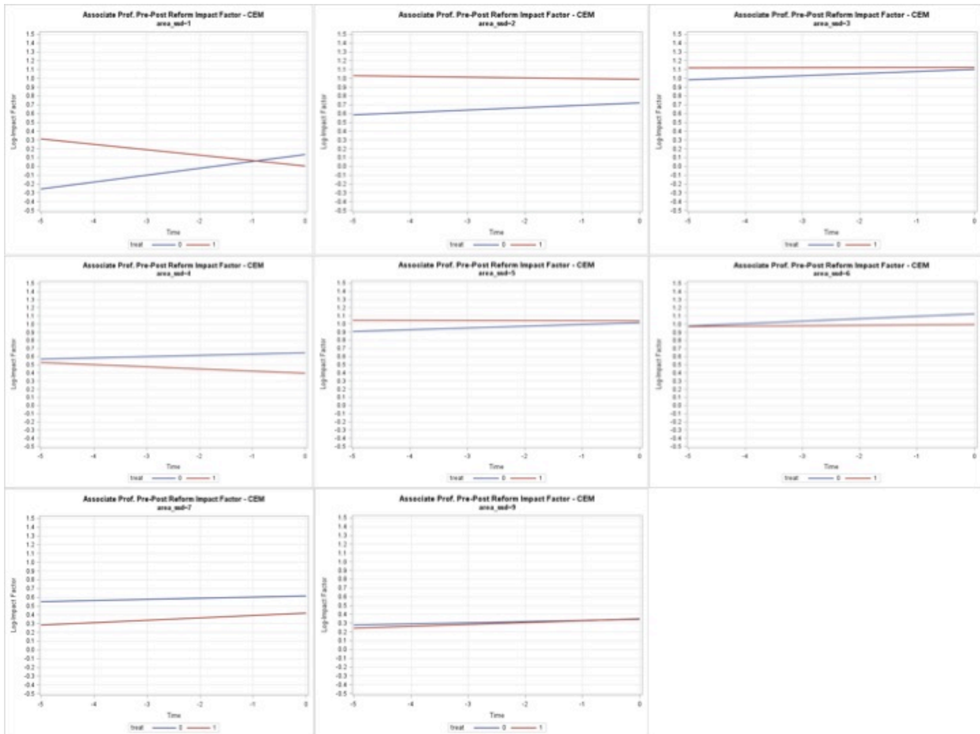


Figure 14: CEM - Research impact growth lines for Associates' by academic discipline, region controls included

4.7 Conclusions

Heterogeneity is endemic in the Italian University system and it is not a surprise to obtain different results across academic disciplines; the story is different for each discipline and it is certainly not possible to generalize the results to the entire Italian Academia. To account for these considerations, as well as data availability issues for international oriented research outcomes, we focus on the hard sciences. We find shrinking individual research productivity (both in terms of slopes and six-year levels) due to local selection mechanisms (with respect to national ones). Lower incentives for publishing on international top-level journals are associated with the decentralization reform.

In sum, the main results are the following:

- Propensity score matching regressions provide evidence of negative effects on slopes only for impact factors while, controlling the balance of local and nationally selected individuals with the non-parametric coarsened exact matching technique, the model stresses how both intercept and slopes are negative and statistically significant for impact of research outcomes on associate and full professors.

- Quartile propensity score matching models show that, in terms of research impact, the negative effect is persistent along the impact-distribution of researchers, while the productivity index (n° of paper on ISI) is significant only for the bottom of the distribution (first quartile) with associated negative effects (this means that top researchers were not influenced by the selection mechanism in their research pipelines);

- For the hard sciences, after-CEM we found negative effects of decentralization on both final outputs levels and slopes for impact factors of both associates and full professors. The results were statistically significant and negative only for slopes the N° of paper on ISI (both for associates and full professors);

- Differentiating by research areas, the negative incentives of local selections six years after selection research levels are statistically significant in the: Math(-), Earth sciences (-), Medicine (-), Veterinary and Agricultural sciences (-); Differentiating by research areas, the negative incentives of local selections on slopes are significant only in Math(-) and Agricultural and Veterinary areas (-);

- Results on slopes are robust for different econometric strategies of balancing between treated and controls [propensity score matching, coarsened exact matching] while results on intercepts (final levels) are statistically significant only with CEM;

- Effects are robust on pre-1994 vs. after-2000 professors (used as controls of the incentive schema due to the possibility of mixed incentives for people selected as associates after 1994 but before 2000 with national procedures, and applying as full professors with the local mechanism).

Otherwise there are several shortcomings that we know affect our estimates:

- Our analysis is focused on the effect of positive/negative incentives of the recruitment reform in Italy (1998) on research outputs, calculated using data of publications on international journals only (due to data availability). Italian language papers and conference proceedings or books are excluded by the information source;

- No data are available on teaching evaluation of Italian academics, so that the effort of the applicant on teaching dimension could not be considered by our models, while this consideration is well-known for its role in academic recruitment as it is one of the primary departments needs (Noser et al. 1996).

- Credible results are available for "Hard sciences" only, due to the higher exposition of these disciplines onto the international research area (collected by ISI). Arts and Humanities and Social Sciences could not be tested due to the inconsistence of available data;

- We have no data on the relative importance of teaching and research as the two main dimensions of academic recruitment (this is one research objective we expect to investigate in the future). The effort of candidates is usually divided on both these activities in the years before each concorso, and it is likely the case that most universities in the last decade judged candidates with respect of both their level of research outcomes (quantity and impact) and teaching.

- The shrinking incentives in the research dimension due to decentralization are certain, yet we know very little about teaching incentives.

Trying to summarize few policy indications from our work we argue that key problems arise from the scarcity of competition between universities in Italy. An insufficient mobility rate of professors within the country (typical of local mechanisms), the predominance of institutional needs (both in research and teaching), and the preeminence of a small number of academic "schools" with respect to others, all have had a negative impact on research growth paths of local recruited researchers.

In the Italian hiring system no penalties were associated with collusive behaviors, and the lack of competitiveness between institutions is endemic in the Italian university system. This is probably due to the "false-autonomy" of universities where salary levels and teaching loads are centrally regulated (by MIUR). No incentives could be driven by the single universities.

In some countries (UK, USA, Australia et cetera), higher quality professors, both in terms of research publication records and teaching abilities, are sought after by universities in a competitive market where salary benefits and teaching load are key levers in the hands of the institutions. In Italy, there are few incentives that encourage candidates to reach high publishing goals, and decentralized recruitment mechanisms reduce these incentives further.

References

1. Abadie, A., Angrist, J. and Imbens, G., (2002) “Instrumental Variables Estimation of Quantile Treatment Effects,” Econometrica, Vol. 70, No. 1, 91-117;

2. Adler R., J. Ewing J., Taylor P., (2009) “Citation statistics (with discussion)”, Statistical Science, 24, 1-28;

3. Allesina, S. (2011), “Measuring nepotism through shared last names: the case of Italian academia”, PLoS one 6(8), e21160;

4. Angelucci, M., G. De Giorgi, M. Rangel and Rasul I. (2010), “Family networks and school enrollment: evidence from a randomized social experiment”, Journal of Public Economics 94(3-4), 197–221;

5. Archambault E., Campbell D., Gingras Y., Larivire V. (2009) “Comparing Bibliometric Statistics Obtained from the Web of Science and Scopus”, Journal of the American Society for Informatin Science and Technology, 60(7);

6. Bakkalbasi N., Bauer K., Glover J.,Wang L., (2006) “Three options for citation tracking: Google Scholar, Scopus and Web of Science”, Biomedical Digital Libraries, 3-7;

7. Bitler, M., Gelbach, J. and Hoynes, H., (2002) “What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments” Working Paper, Department of Economics, University of Maryland;

8. Checchi, D. (1999), “Tenure. an appraisal of a national selection process for associate professorship”, Giornale degli Economisti 58(2), 137–181; Durante, R., Labartino G., Perotti R., and Tabellini G. (2009), “Academic dynasties”, Bocconi University Working Paper;

9. Cochran W.G., (1968), "The Effectiveness of Adjustment by Subclassification in Removing Bias in Observational Studies," Biometries, 24, 205-213;

10. De Battisti, F., Salini, S., (2011) “Robust analysis of bibliometric data”, Departmental Working papers 2011-36, Department of Economics, Management and Quantitative Methods at University of Milan; D’Agostino, R. B., (1998) “Tutorial in biostatistics. Propensity score methods for bias

reduction in the comparison of a treatment to a non-randomized control group", Statistics in Medicine, Vol. 17, 2265-2281;

11. Falagas M.E., Pitsouni E. I., Malietzis G. A.,Pappas G., (2008) "Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strenghts and weaknesses", The FASEB Journal, 22, 338-342;

12. Fox, M. F. (1983), "Publication Productivity Among Scientists: A Critical Review", Social Studies of Science, 13, 283-305;

13. Garibaldi P., (2006), "Personnel Economics in Imperfect Labour Markets2, Oxford University Press, chapter .8;

14. Heckman, J., (1979) "Sample Selection Bias as a Specification Error", Econometrica, Vol. 47, No. 1, pp. 153-161;

15. Hedeker, D., and Gibbons, R.D., (1997) "Application of Random-Effects Pattern-Mixture Models for Missing Data in Longitudinal Studies", Psychological Methods, Vol. 2, No. 1, 64-78;

16. Howell, D.C., (2008) "Mixed models for repeated (longitudinal) data";

17. Iacus, Stefano M., and Giuseppe Porro., (2007) "Missing data imputation, matching and other applications of random recursive partitioning", Computational Statistics and Data Analysis, 52:773–89;

18. Iacus, S. M., King G., and Porro G., (2009) "CEM: Coarsened Exact Matching Software", Journal of Statistical Software 30(9);

19. Iacus, S. M., King G., and Porro G., (2011a) "Replication data for: Causal inference without balance checking: Coarsened Exact Matching", Murray Research Archive;.

20. Iacus, S. M., King G., and Porro G., (2011b) "Multivariate matching methods that are Monotonic Imbalance Bounding", Journal of the American Statistical Association;

21. Imbens, G., and Wooldridge, J.M., (2008) "Recent developments in the Econometric of Program Evaluation", IZA DP No. 3640;

22. Lazear, E. (1995), "Personnel economics", MIT, Chapter 3;

23. Klavan and Boyak, (2007) “Is there a Convergent Structure of Science? A Comparison of Maps using the ISI and Scopus Databases1”, 11th International Conference of the International Society for Scientometrics and Informetrics meeting in Madrid, Spain, June 2007;

24. Kleinman, K., Horton, N.J., (2009) “SAS and R: Data Management, Statistical Analysis, and Graphics”, Ed. Chapman and Hall/CRC;

25. Levin S., Stephan P. (1991), “Research Productivity Over The Life Cycle: Evidence for Academic Scientists”, The American Economic Review, Vol. 81, Issue 1, 114-132;

26. Mielke and Berry (2007), “Permutation methods: A distance function approach”, New York: Springer;

27. Myers, J. A. and Louis, T. A., (2010) “Regression adjustment and stratification by propensity score in treatment effect estimation”, Johns Hopkins University, Dept. of Biostatistics Working Papers. Working Paper 203

28. Norris M., Oppenheim C., (2007) “Comparing alternatives to the Web of Science for coverage of the social sciences literature”, Journal of Infometrics, 1,161-169;

29. Noser T., Manakyan H. and Tanner J. (1996), “Research Productivity and Perceived Teaching Effectiveness: A Survey of Economics Faculty”, Research in Higher Education, Vol. 37 No. 3

30. Overall, J. E., Ahn, C., Shivakumar, C., & Kalburgi, Y. (1999) “Problematic formulations of SAS Proc.Mixed models for repeated measurements”, Journal of Biopharmaceutical Statistics, 9, 189-216;

31. Rosenbaum, P. R., and Rubin, D. B., (1983) “The central role of the propensity score in observational studies for causal effects”, Biometrika, Vol. 70, 41-55;

32. Seglen P.O. (1997), “Why the impact factor should not be used for evaluating research” British Journal of Medicine, 314, 498-502;

33. Shah B.R., Laupacis A., Hux J.E., Austin P.C. (2005) “Propensity score methods give similar results to traditional regression modeling in observational studies: A systematic review”, Journal of Clinical Epidemiology, 58:550–55;

34. Stuart, E. A.; Cole, S. R.; Bradshaw, C. P.; and Leaf, P. J., (2010) "The use of propensity score to assess the generalizability of results from randomized trials", Johns Hopkins University, Dept. of Biostatistics Working Papers, n° 210

35. Weitzen S., Lapane K.L., Toledano A.Y., Hume A.L., Mor V., (2005) "Weaknesses of goodness-of-fit tests for evaluating propensity score models: The case of the omitted confounder", Pharmacoepidemiolgy and Drug Safety,14:227–238;