What do you want to be when you grow up? Local institutional quality and the choice of the fields of study in Italy (2004-2007)

Nifo, Annamaria and Scalera, Domenico and Vecchione, Gaetano

8 March 2016

"What do you want to be when you grow up?". Local institutional quality and the choice of the fields of study in Italy (2004-2007)

Annamaria Nifo[°] **Domenico Scalera**[°] **Gaetano Vecchione**[◊]

[°] Department of Law and Economics (DEMM) – University of Sannio – Benevento, Italy; nifo@unisannio.it
[°] Department of Law and Economics (DEMM) – University of Sannio – Benevento, Italy; scalera@unisannio.it
[◊] Department of Political Science – Second University of Naples – Caserta, Italy; gaetano.vecchione@unina2.it

Abstract

Students' choices about post-secondary fields of study vary widely across space and time, due to many psychological, social and economic motivations. Regarding these latter, the most important role in steering students' options has been often ascribed to expected returns from different occupations. This paper emphasizes in particular the link between local institutional quality, the reward structure and students' preferences. Based on a sample of 80,996 students graduated in Italy in 2004 and 2007, our econometric investigation, controlling for both individual characteristics (gender, residence, family background, high school track) and geographical variables (per capita GDP, industrial specialization), finds that in the choice of the field of study institutional quality definitely matters.

Keywords: Institutional quality; Fields of study; Regions; Italy

JEL Classification: D02; E02; I21

1. Introduction

Students' choices about post-secondary fields of study vary widely across space and time. According to UNESCO Institute for Statistics (2014), in 2011-2012 within European Union, the share of graduates in engineering ranged between 8.4% of Malta to 20.1% of Finland, while those obtaining a degree in law, social sciences or business were just 23.4% of total graduates in Germany but more than 54% in Bulgaria and Romania. Similar differences can be found elsewhere as well: in America in the same years, the proportion of graduates in law, social sciences or business largely varied between Cuba (24.2%) and Chile (29.4%) on one side, and Brazil (41.0%) and Colombia (48.5%) on the other. Concerning differences over time, the same source reports that between 1999-2000 and 2011-12 the share of graduated in law, social sciences or business has undergone substantial increases in many advanced countries, with an average rise of about 4 percentage points and a growth of 9.3% in Australia. Even more notably, with reference to a large sample of US students, Goldin and Katz (2008) show that the shares of those completing medical and business school respectively decreased by 19% and increased by 111% between 1972 and 1992.

The reasons behind this wide heterogeneity in students' preferences have been long investigated. To justify so divergent attitudes and ample discrepancies among different countries and times, many psychological, social and economic motivations (some of which are recalled below in the review of Section 2) have been set forth. Regarding in particular economic drives, the most important role in steering students' options has been often ascribed to expected returns from different occupations, while a number of contributions have focused on other microeconomic aspects of the choice (individual ability and tastes, the cost of education, the presence of financial constraints, etc.). On the other hand, it has been also recognized that often macroeconomic structural factors matter as well: for example, productive specialization, technology, markets' competition and institutional factors have been deemed to be relevant in driving students' preferences, to the extent that they affect the reward structure and shape differences in careers' relative expected profitability.

This paper adopts an approach emphasizing in particular the link between local institutional quality, the reward structure and students' preferences. As recalled at length in Section 2, a wide literature has argued that bad institutions encourage rent-seeking, thus pushing talent toward either occupations meant to defend from rent-seeking behaviors (mainly legal and socio-political professions) or rent-seeking activities themselves. Conversely, when institutions are strong, careers devoted to shelter from rent-seeking are less needed and less profitable: less needed because a smaller share of production is at stake, and less profitable because rent-seeking is less successful (Acemoglu, 1995). Building on several previous studies, we consider choices for science, engineering and economics degrees or alternatively for law or socio-political studies respectively as proxies for an option for productive activities or alternatively in favour of rent-seeking (or protection from rent-seeking). This latter choice has to be seen merely as a rational response to relatively high rewards for legal and socio-political professions supplied by a context of poor institutional quality, so that obviously no negative moral judgment is involved.

Addressing the issue of the choice of post-secondary education in Italy has a strong motivation in the evidence of pretty large differences in students' preferences among different regions in the country. For example, according to the representative sample of ISTAT (2009; 2013), that we use in this paper, the share of students graduated in Business, Science (including Chemistry), Engineering and Architecture is around 39% in the North-Western (Lombardy, Piedmont, Liguria and Valle d'Aosta), and North-Eastern (Veneto, Emilia Romagna, Trentino Alto Adige and Friuli Venezia Giulia) regions, 34% in the regions of the Centre (Toscana, Lazio, Marche and Umbria) and 33% in the South (Abruzzo, Campania, Apulia, Molise, Basilicata, Calabria, Sicily and Sardinia). Conversely, students of Law and Socio-political disciplines account respectively for 15% in the North-West, 14% in the North-East, 17% in the Centre and almost 20% in the South. More recent data made available by Almalaurea (2013), collecting data on graduates of 64 (out of 85) Italian universities, confirm the extent of differences: degrees in Science, Engineering, Architecture and Business account for 40.7% of total graduates in the North, but only

for 35.2% in the South, whereas degrees in law and socio-political studies are obtained by less than 15% of students in the North versus more than 19% in the Southern regions.

Explaining large differences at a sub-national is especially puzzling, considering that, at least in developed countries, economic conditions, production structures and career rewards are usually quite homogeneous across regions. In the case of Italy however the evidence of significant diversity in students' preferences is less surprising, in the light of the large disparities occurring in a number of economic and social indicators across different regions of the country (Malanima e Zamagni, 2010; Giannola *et al.*, 2015) which testify the multifaceted nature of the Southern lag. In particular, this paper focuses on the heterogeneity in the institutional quality of Italian regions as a possible determinant of students' choices on post-secondary fields of study. To this purpose, the Institutional Quality Index (IQI) constructed by Nifo and Vecchione (2014) is employed. IQI evaluates local institutional quality for Italian regions and provinces as a synthetic indicator derived by 24 elementary indexes grouped into five categories (voice and accountability, government effectiveness, regulatory quality, rule of law, corruption), in the fashion of the World Governance Indicator (WGI) proposed by Kaufmann *et al.* (2010).

The econometric investigation is conducted on a sample of 47,300 students graduated in Italy in 2004. Estimation methods are Probit, Logit and Multinomial Logit models. Given the hierarchical nature of data, for robustness purposes, estimation is also made through a multilevel analysis considering both individuals (at a lower level) and regions as aggregate units at a higher level. Our findings are definitely robust and basically confirm the hypothesis that local institutional quality plays a significant role in determining students' choice. As a matter of fact, once controlled for students' individual characteristics (gender, family conditions, migratory status, secondary school attainments), the reputation of different universities for each field of study, the economic condition of students' province of origin (per-capita GDP and industrial intensity), we find that institutional quality is relevant to the choice of the field of study, with relatively high marginal effects. This outcome holds in particular for the "Rule of law" sub-index, emphasizing that

institutional contexts characterised by a relatively high incidence of crime, poor law enforcement and low effectiveness in the administration of justice spur a higher demand for protection from rent-seeking and make it more profitable to acquire education in the fields of law and social and political studies.

Following this introduction, Section 2 provides a short review of the literature on the issue. Section 3 is devoted to the econometric investigation. In particular, in Section 3.1 estimation strategy, methods, and explanatory variables are presented; in Section 3.2 the dataset and some descriptive statistics are shown; Section 3.3 displays results and comments on them. The main conclusions of the paper are collected in Section 4.

2. Related literature

The choice of the post-secondary field of study has drawn considerable attention by economists through time and given rise to a large number of both theoretical and empirical papers. This literature can be considered as a segment of the much wider literature flourished on human capital and the decision to invest in education since the seminal work of Becker (1964), just like the issue of selecting the field of study for a given education level (i.e., the educational horizontal choice) may be seen as an aspect of a more general option about the amount of education to acquire (i.e. the vertical choice among primary, secondary or post-secondary education).

The specific literature on the choice of the field of study has developed by adopting at least three possible different approaches. The prevailing line of research has privileged a mostly empirical microeconomic viewpoint, by focusing on expected future earnings and returns to education as the key variables in shaping students' options (Kirkebøen *et al.*, 2014; Beffy *et al.*, 2012; Boudarbat, 2008; Berger, 1988). In a nutshell, according to this approach, different occupations offer very different payoffs, and individuals tend to choose careers in which they have comparative and absolute advantage. Of course, even this perspective does not deny that many other variables may be relevant to the choice (Wiswall and Zafar, 2015; Fiorito and Dauffenbach, 1982).

Indeed, selecting the field of study for a given education level is much more complex than simply determining the number of years of education, and this choice may be reasonably affected by a number of non strictly economic factors. As Altonji *et al.* (2011) point out, beside the purely financial returns to education, which include expected wages, net tuition costs, and the discount rate, the students' choices also "depend on ability, the prior stock of knowledge, and tastes for education". In addition, individual risk aversion matters for several reasons, for "individuals care about uncertainty in ability and the effect of this uncertainty on alternative degrees or courses of study, possibly avoiding ones where ability might matter a lot", and also because of "the possible volatility of returns to education (…) and the microeconomic uncertainty in wages within each sector".

However, as the dominant approach does not seem to fully account for all possible psychological and social motivations behind preferences for different types of education, another strand of the literature, building on the seminal work of Akerlof and Kranton (2002), emphasize the concept of social identity as a major determinant in the choice of careers and fields of education. In this view, purely economic factors like expected net economic rewards are little suitable to explain why highly able individuals choose so frequently high-skilled low-paid occupations. According to Humlum *et al.* (2012) "individuals (…) think of themselves and others in terms of social categories. In addition to the standard pecuniary payoff, the utility function incorporates a non-pecuniary payoff associated with a person's identity, which depends on the agent's choice of social category and on how well the agent's characteristics and actions correspond to the ideal of that social category". Careers and fields of study are therefore somehow connected to the type of identity that each student assigns to herself or himself: so, individualistic doers, like for example most of "nerds", are more incline to studies like Business, Law and Economics, while characters associated to a more social identity would rather propend for Teaching, Humanities and Arts. This setting shows to be particularly fruitful in interpreting gender differences in access to careers and fields of study and explaining why, even in the most advanced countries, women "remain substantially under

represented across a range of technical and scientific fields" (Rosenbloom *et al.*, 2008). Also, other personal and social features such as marital status, race and the family cultural background are found to matter significantly in students' choice of the field of studies (Montmarquette *et al.*, 2002; Boudarbat and Montmarquette, 2009).

While both these approaches focus on individual determinants of the choice of the field of study, neither of them explicitly considers how the macroeconomic environment (for example, technological structure, wealth distribution, institutional quality) might affect that choice. Conversely, the impact of these latter factors on returns to different types of education and the profitability of different occupations has been investigated by another strand of the literature. For example, Banerjee and Newman (1993) show the interplay between agents' occupational decisions, production technology and the distribution of wealth; Acemoglu (1995) argues that the reward structure of entrepreneurial careers and rent seeking activities depends on technological factors as well as the extent of the negative externalities exerted by rent seeking on productive activities; in the same vein, Mehlum *et al.* (2003) claim that at low levels of development, predation is often more attractive than at higher levels of development, so that a larger share of agents prefers to be engaged in criminal rather than productive activities, thus pushing the economy in a vicious cycle of poverty and predation.

Concerning specifically institutions, the role of well defined property rights, little corruption, definite and secure rule of law is documented by a broad literature, according to which institutional quality exerts an impact on both the accumulation of human capital and the choice of the kind of education. A positive association between institutional quality and human capital accumulation is found for example by Pecorino (1992) and Hall and Jones (1999). According to Pritchett (2006), the effect of institutions on the investments in human capital is more qualitative than quantitative, as they shape the structure of rewards and affect the choice of careers, thus determining how really beneficial is human capital for growth. The idea that different institutions supply different individual incentives relevant to the choice of the field of study is also developed

by Natkhov and Polishchuk (2013), who directly connect institutional quality to the choice of college majors in Poland and Ukraine. Building on the evidence of large differences in the shares of students pursuing degrees in science and technology or in legal studies, these authors argue that those differences reflect the uneven quality of institutions in the two countries: "Stronger institutions make young Poles confident in the ability to earn good rates of returns to their skills, knowledge, and innovations in modern technologies, whereas young Ukrainians believe that law degrees would better equip them for an institutional environment where the rule of law is feeble and corruption and rent-seeking reign supreme".

Treating a degree in science or engineering as a proxy for the choice of an occupation in productive activities, and instead an option in favour of a career in legal studies as a preference for activities of (protection from) rent-seeking dates back at least to Tollison (1982) and has been widely employed in the literature. Magee *et al.* (1989) and Murphy *et al.* (1991) are among the first to use the number of lawyers or the relative college enrolment rates in law and in engineering in a growth equation as proxies for the society attitude to rent-seeking or productive entrepreneurship. More recently, many others have highlighted on one side the role of good institutions in favoring the demand of knowledge and technical abilities and then higher rewards for science and engineering careers (Levchenko, 2007; Nunn, 2007); on the other the incentive to legal professions supplied by a context of poor institutional quality. On this latter point, several contributions have clarified that bad institutions underpin the appeal of legal studies not only by increasing chances and profitability of rent-seeking but also by raising the demand for protection against rent-seekers and other threats associated to poor institutional quality (Arruñada, 2007; Dezalay and Garth,1997; Cumming and Johan, 2006)[1]. Finally, like for legal studies, even the preference for degrees in political and social studies may be treated as an indicator of propensity to careers in (protection from) rent seeking: indeed many authors, from different viewpoints, have emphasized the role of

[1] Magee (2010) distinguishes between facilitative legal actions (protecting and increasing wealth) and redistributive actions bringing about opposite effects.

politicians and bureaucrats in rent-seeking activities, intermediation and protection from rent-seeking, lobbying and bribes (e.g. Krueger, 1974; Mushtaq and Jomo, 2000; Keefer and Knack, 2007).

To conclude, we like recalling how the relationship among institutions, the agents' endowment of skills and economic growth is acutely summarized by North (1992): "if the institutional framework made the highest pay-offs for organizations piracy, then organizational success and survival dictated that learning would take the form of being better pirates. If on the other hand productivity raising activities had the highest pay-off then the economy would grow". In the same vein, Murphy *et al.* (1991) point out that the quality of institutions is decisive in steering choices toward productive activities or rent seeking because "Countries with poorly defined property rights attract talent into rent seeking, since success at redefining these property rights brings high rewards. Rent seeking pays because a lot of wealth is up for grabs". Some implications of this intuition are displayed in Appendix 1 of this paper, where we sketch a simple formal framework built on Acemoglu (1995), showing how institutional quality, represented by the share of production subject to be seized by rent seekers, may impact on incentives and determine the choice among careers in production, rent seeking and protection from rent seeking.

3. The empirical investigation

This section is devoted to provide evidence about the reasons driving the choice of post-secondary field of studies in Italy and in particular to single out the role of local institutional quality in determining this choice. To perform this task, we carry out an econometric analysis, where the students' option is the dependent variable. Estimation strategy and methods, and the motivations behind our choice of regressors are presented in Section 3.1. The dataset we employ, together with some descriptive statistics are illustrated in Section 3.2. Finally, Section 3.3 deals with the presentation and discussion of the results of our investigation.

3.1. Estimation strategy and methods

Since the students' choice about the field of tertiary studies can be clearly seen as a qualitative variable, methods eligible for evaluating the role of possible explanatory factors of this choice necessarily belong to the family of the limited Dependent Variable models. For this reason, the following analysis is conducted by employing Logit and Probit models for binary response, and then also a Multinomial Logit model. Consistently, in the case of Logit and Probit estimations, the students' choice is modeled as a binary option between *Productive* and *Defending* fields of studies and regressors are assumed to impact on the probability that a student makes each choice. In accordance with the approach followed by most of the literature on the relationship between institutions and human capital accumulation recalled in Section 2, we include studies in Science, Chemistry and Pharmacology, Geology, Biology, Engineering, Architecture, Economics and Statistics in the group of *Productive*; the group of *Defending* is made up of Law and Political and social studies; the group of "Other fields of study" comprises Literature, Language, Teaching, Psychology, Physical education, Defence and Security. Finally, Medical studies are not included in any group because of the peculiar restrictions legally imposed on the number of students enrolled in this field of study (*numerus clausus*), which makes it unsuitable to be used in the econometric investigation.

More precisely, the model we estimate is

$$Prob(Y = 1|\boldsymbol{X}) = G(\beta_0 + \boldsymbol{X\beta}) \qquad [1]$$

where Y is the students' choice, conventionally assuming the value 1 if a student prefers a *Productive* field of study and 0 in case that a *Defending* one is chosen; $\boldsymbol{X}$ denotes the vector of explanatory variables, β are parameters and G is alternatively the logistic function in the Logit model and the standard normal cumulative distribution function in the Probit model.

Secondly, we check for the robustness of our results by also considering another theoretical framework, allowing for a third alternative students' choice, i.e. "Other fields of studies". In this case, we resort to a Multinomial Logit (MNL) model, with response probabilities

$$Prob(Y = j|\boldsymbol{X}) = \frac{exp(\boldsymbol{X\beta}_j)}{\sum_{j=0}^{2} exp(\boldsymbol{X\beta}_j)}, \qquad j = 0, 1, 2 \qquad [2]$$

where again Y is the choice, assuming now three possible values, i.e. 0, 1 or 2 if the student respectively chooses *Defending*, *Productive* or *Other fields of studies* and the other symbols are the same as above. The MNL model seems a good candidate to carry out our analysis for several reasons: the dependent variable of our regression equation includes multiple non-ordered responses; the dataset contains many strictly individual data for which the utilization of MNL is suggested by many econometricians (Cameron and Trivedi, 2010); it has been used by other studies on the determinants of the fields of studies (Nguyen and Taylor, 2003; Boudarbat, 2008).

A further robustness check consists in re-estimating the effect of explanatory variables on students' choice by adopting a multilevel rather than a single level model (Hox, 2010; Luke, 2004; Goldstein, 2003; Kreft and de DeLeeuw, 1998). Multilevel models are specifically designed to deal with statistical data displaying a clustered construction, like in our case, where individual students may be considered lower level units, and administrative regions higher level aggregate units. Since it is reasonable to assume that students residing in the same region share some common features and are more similar to each other than students living elsewhere, the assumption of independence of errors is likely to be violated. By considering group-level variance through the inclusion of random coefficients, the multilevel approach allows to control for spatial dependence and correct the measurement of standard errors, thereby ensuring efficient estimates[2].

The vector $\boldsymbol{X}$ includes a number of different regressors according to the various specifications that we consider. Explanatory variables account for different aspects of the choice, conveying information on i) students' individual characteristics, such as gender and parents' education and occupation; ii) students' secondary education, i.e. High School grade and High

[2] Estimation is made by using STATA and assuming that regression parameters β of equation [1] are region specific, i.e. β_k. Modeling these latter to allow for random regional components $\beta_{ik} = \gamma_{i0} + e_{ik}$, where i and k are indexes respectively for parameters and regions, individual observations $Y_{hk} = \beta_{0k} + \sum_{i=1} \beta_{ik} X_{ihk} + \varepsilon_{hk}$ become $Y_{hk} = \gamma_{00} + \sum_{i=1} \gamma_{i0} X_{ihk} + (e_{0k} + \sum_{i=1} e_{ik} X_{ihk} + \varepsilon_{hk})$, where on the right hand side the hierarchical structure of the data is represented by a deterministic part (the first addends) and a stochastic part (the last addends in brackets).

School type; iii) the quality of courses taught at different universities for each field of study; iv) the economic condition of students' province of origin, specifically in terms of per-capita GDP and an index of industrial intensity; v) institutional quality of students' region of origin, considered both in terms of the overall value of IQI and the value of single specific IQI dimensions. The employed explanatory variables seem to be little exposed to the possibility of reverse causality, since they concern either individual features which are determined before the student's choice or aggregate variables which is quite unrealistic to suppose to be affected by students' options (if not in the very long run). Table 1 reports the complete list of dependent and explanatory variables used in the econometric investigation.

Our choice of covariates is common to many other papers on the choice of fields of study. Nevertheless, in what follows we motivate the insertion of each explanatory variable. Concerning individual characteristics, we first take into account gender, which a very wide literature argues to be relevant to the choice of the field of study (for a survey, see for example Xie and Shauman, 2003 and Goldin, 2006). It is well known that a remarkable underrepresentation of women is observed in Science, Technology, Engineering and Mathematics (STEM) professions (House of Commons, 2014; European Commission, 2009), scientific academic careers (European Commission, 2009), and participation in corporate boards (UK Government, 2011) in many countries, although to a different extent (Jacobsen, 2007). The reasons for this phenomenon are several: without denying the possibility of differences in preferences between genders, segregation and discrimination are certainly important as well. Segregation, i.e. the tendency for men and women to be employed in different occupations (Siltanen *et al*., 1995) due to cultural and sociological reasons, is widely recognized by the literature as a persistent worldwide phenomenon (Charles and Grusky, 2004) and a source of gender wage differentials. Instead, gender discrimination occurs when wage gaps are not completely explained by characteristics related to productivity such as skills, abilities, education, experience. According to most of the empirical literature, the discrimination component of the gender wage gap, although decreasing over time, is still quite high (about 20%) and tends to decline

much slower than the overall differential (Weichselbaumer and Winter-Ebmer, 2005). In terms of choice of field of studies, discriminatory wage differentials may affect women's options: as discrimination is in absolute terms more harmful for more remunerative occupations, it reduces women's reward for technical and scientific careers. Recent literature (Buser *et al*., 2014) has also pointed out that gender differences in educational choices may be explained by a lower tolerance for competition (rather than lower abilities) among women, deterring them from entry into more challenging and remunerative fields of study.

Other individual factors affecting students' educational choices are related to family cultural, social and economic conditions (Haveman and Wolfe, 1995; Esping-Andersen 2004; Goodman *et al.*, 2015). Since more educated parents tend to encourage choices in favor of more profitable and prestigious careers, a limited intergenerational mobility in schooling and educational attainment has been observed by many authors (Dustmann, 2004; Corak, 2006; Hertz et al., 2007). Others have associated the effect of familiar income and wealth on educational choices to the existence of borrowing constraints (Acemoglu and Pischke, 2001; Carneiro and Heckman, 2002). In our case, the absence of data on family income or wealth, induces to include parents' education and occupation among regressors not only as intrinsically relevant variables, but also as proxies of familiar economic conditions. However, while the expected effect of family income and wealth on the *amount* of education to acquire is likely to be positive, the impact on the *kind* of tertiary studies is much less obvious. Living in an environment more affluent and more familiar with business (as in the case of entrepreneurs) may give more awareness of the importance and gains associated to a "defending" occupation. Also, if higher income is associated to less risk aversion, students coming from richer families should show higher propensity to pick studies and occupations where ability (not thoroughly known *ex-ante*) and fortune might matter more. So, for example, if becoming a successful lawyer is a highly uncertain occurrence, a higher familiar income and a larger professional network of parents may help to choose studies in Law.

Secondary education type and grade are also considered to account for possible effects of students' background and abilities on the choice of tertiary education. The impact of the type of secondary school on the choice of majors is investigated for example by Checchi and Flabbi (2013), who focus on the effects of early or late tracking scholastic systems and Boeri *et al*. (2015) within an analysis of the gender wage gap. The importance of high school grades hinges on the idea that individuals with higher ability can harvest the largest benefits from investment in tertiary education (Averett and Burton, 1996; Hilmer, 1998). In recent papers, Wang (2013) finds a positive correlation between high school grade, especially high school math achievement, and the choice for a STEM major, while Natkhov and Polishchuk (2013) consider grades obtained at secondary school as an explanatory variable for students' choice of tertiary education because of the possible different impact of institutions on the allocation of talent according to students' abilities.

The quality of colleges is also considered an important driver of students' enrollment choices (Long, 2004; Luca & Smith, 2013). Especially for those little inclined to move outside the region or the area of origin, the relative quality of courses in scientific or legal studies may significantly affect the choice for each field of study. To capture this "supply-side" effect, we build up a regional Relative Evaluation of the Engineering Faculty (REEF) index. In order to construct the regional REEF index, we consider for each region only universities where (at least) one tertiary education program in Engineering and one in Law are taught. Then for each university, the normalized ratio of ANVUR (Italian National Agency for the Evaluation of University and Research) 2004-10 assessments of the Engineering and Law faculties is considered[3]. Finally, a regional average is calculated based on universities' geographical location. If the quality of faculties

[3] The ratio is made by the numerical evaluations (on a range from 0 to 1) reported at the websites http://www.anvur.org/rapporto/files/Area09/VQR2004-2010_Area09_Tabelle.pdf (Table 4.7.a) and http://www.anvur.org/rapporto/files/Area12/VQR2004-2010_Area12_Tabelle.pdf (Table 3.1), respectively for Engineering and Law faculties. Since throughout the country Engineering gets an average grade of 0,72 and Law 0,50, ratios are normalized by dividing by 72/50, so that for Italy REEF=1.

matters, we expect that regions with a high relative quality of Engineering (Law) display higher shares of students enrolled in Engineering (Law)[4].

The fourth kind of variables we consider accounts for the economic condition of students' province of origin. In particular, we include among regressors per-capita GDP and a provincial index of industrial intensity (i.e. the number of industrial employees over population). We expect that these variables contain information on structural economic factors possibly interacting with careers' profitability. Indeed, the presence of a developed industrial environment might make more profitable for students a choice for *Productive* studies, by engendering a strong and sustained demand for engineers and a workforce endowed with adequate technical skills[5].

Finally, we employ one or more indicators of regional institutional quality as explanatory variables for the choice of the field of study. The presence of institutions among regressors is what specifically characterizes this paper and its motivations have been expressed in detail in Section 2. In addition, Appendix 1 sketches a simple theoretical framework, inspired to the models of Murphy *et al.* (1991) and Acemoglu (1995), aiming at demonstrating how and why institutional quality may affect the reward structure of professions, the profitability of *Defending* occupations and ultimately the choice of students.

3.2. Data and descriptive statistics

Our investigation is based on a unique dataset provided by ISTAT (Italian National Institute of Statistics) which periodically surveys graduates' employability three or four years after graduation. In particular, we use the seventh and the eight wave of the ISTAT survey, reporting information on respectively a sample of 47,300 students graduated in 2004 and interviewed in 2007, and a sample

[4] Most of the extant literature finds that the effect of (changes in) university rankings on the number of applications received is statistically significant even if quantitatively not large (see for example, Broecke, 2015 on UK; Clarke, 2007 on USA and Muller and Rockerbie, 2005 about Canada).

[5] This connection has been documented in particular for the establishment of new large industrial plants. For Southern Italy, Cersosimo and Viesti (2013) have recently examined six case studies, showing the existence of a close relationship between high tech industrial development and the rise in the number of engineers, physicists and technicians graduated at local universities.

of 33,696 students graduated in 2007 and interviewed in 2011[6] (ISTAT, 2009 and 2013). The survey includes data on graduates' individual characteristics (gender, age, residence, family background, high school track) and educational choice (field of study and geographic location of the university). To supplement this dataset, we also use provincial data for per capita GDP and industrial specialization (ISTAT, 2010 and 2011). Finally, we exploit the information on local institutional quality in Italian regions contained in the Institutional Quality Index constructed by Nifo and Vecchione (2014), and in each of the five indicators constituting IQI, i.e. voice and accountability, government effectiveness, regulatory quality, rule of law and corruption (in turn made up by 24 elementary indexes[7]).

Preliminary descriptive statistics on our sample and the population are delivered by Tables 1 to 4. Table 1 displays the list of the variables with minima, maxima and averages for each variable. Table 2 shows the distribution among different fields of study of Italian graduates in 2004 and 2007. The left panel of Table 2 shows that economics and statistics, engineering, political and social studies, medical studies and law are the most popular subjects in university students' choices, accounting together for about 60% of total Italian graduates. Scientific fields (science, chemistry and pharmacology, geo-biological studies) are preferred by a little less than 10% of students; arts and humanities (architecture, literature, language, psychology and teaching) account for about 26.6% (2004) or 28.5% (2007) of population, while students getting a degree in agricultural studies, physical education or defence and security amount to less than 3.5% of population. The right panel of Table 2 rearranges the data, by grouping fields of studies into categories which are more consistent with our analytical purposes, aiming at contrasting *Productive* and *Defending* (from rent-seeking) careers. When aggregating fields of study in this fashion, we get the groups of *Productive*

[6] More precisely, for the eight wave the number of interviewed graduated amounts to 62,000, i.e. about 20.6% of the entire population of 300,338 individuals against a share of 18.2% of interviewed for the seventh wave, out of 260,070 graduates. However, since the eight wave includes variables carrying sensitive personal information, reasons of privacy justify the reduction of the available dataset. Nevertheless, the reduced sample is statistically representative. The survey is conducted through a detailed questionnaire administered with the Computer Assisted Telephone Interviewing (CATI) technique.

[7] Details on elementary indexes and the data sources are given in Table 1 and Appendix 2.

(Science, Chemistry and pharmacology, Geo-biological studies, Engineering, Architecture and Economics and statistics), *Defending* (Law and Political and social studies) and *Others* (Agricultural studies, Literature, Language, Teaching, Psychology, Physical education and Defence)[8]. *Productive* fields of study show a relative majority of graduates (more than 40%); about 22% of students get a degree in *Defending* fields of study; 11% in Medical studies and the remaining 25% to 27% of graduates are devoted to other studies.

Tables 3 and 4 reconsider our data by geographic area in order to single out possible differences in the distribution among groups of fields of studies. To this purpose, according to the usual partition of national territory, Italy is split into four macro-regions, North-West, North-East, Centre and South. As anticipated in the introduction, Northern regions show a relative prevalence of graduates in scientific, engineering and business fields; conversely, Southern regions display larger shares of graduates in legal and socio-political studies, while central regions are in an intermediate condition. According to the figures reported in Table 3, the largest differences concern Law (3.63 percentage points higher in the South than in North-West) and Engineering (1.44 percentage points higher in North-East than South). However, in relative terms, i.e. dividing the difference shown in Table 3 by the share of Northern graduated in the field of study, the deepest gap is for Science (-43% in the South with respect to the North-East). It is also worthwhile to notice the peculiar distribution of preferences of people graduating abroad: in this case, the shares of Law, Medical Studies and Teaching drastically shrink, while the ones of Business, Engineering and especially Science and Chemistry are considerably higher than for those graduating at domestic universities.

Table 4 rearranges data in a way more consistent to the econometric investigation. Medical studies, together with missing responses and degrees obtained abroad, are left out; fields of study are grouped into *Productive*, *Defending* and *Others*, and distinct figures are provided for each wave (i.e., wave VII: graduated in 2004 and wave VIII: graduated in 2007). This makes differences

[8] Notably, Medical studies are not included in either group because of the peculiar restrictions legally imposed on the number of students enrolled in this field of study (*numerus clausus*). For the same reason, Medical studies are also excluded by the following econometric investigation.

across macro-regions even clearer: the share of students choosing the group of *Productive* in the South is well lower than in other areas (up to 8 percentage points less in 2004, and 7 in 2007) whereas graduated in the group of *Defending* are relatively more numerous (almost 7 percentage points more than in the North-East). Differences in the residual group are less sharp; fields of studies belonging to this group show a decline over time in all macro-regions.

3.3. Results

The results of the econometric investigation are summarized in Tables 5 to 8 for binary response models (Probit in Tables 5 and 6; Logit in Tables 7 and 8), and Table 9 for Multinomial Logit. Tables 10 and 11 report the outcome of multilevel analysis. The interpretation of figures is straightforward: since zero value is associated to the preference for a *Defending* field of study, the sign of estimated coefficients (Coeff) corresponds to the sign of the impact of each regressor on the probability that the alternative choice (*Productive* for Tables 5 to 8, 10 and 11; *Productive* or *Others* for Table 9) is made. Marginal Effects at the Means (MEMs) estimate the quantitative magnitude of such impact, i.e. the effect of a unit change in the value of an explanatory variable on the probability that the choice alternative to the null one is made.

In each table, columns (1) to (5) collect the outcome of regressions run on five different specifications. In Tables 5 to 8 both coefficients and marginal effects are shown. In Tables 5, 7, 9 and 10, the baseline specification, with only familiar and educational background taken into account, is described in column (1). The following specification (column 2) includes the indicator of education quality. A provincial income indicator is then included in column (3). The final specifications comprising institutional quality are described by figures reported in columns (4) and (5), respectively without or with the index of industrial intensity. Tables 6, 8 and 11, report the results of specifications with all explanatory variables, replacing the overall IQI with one of the five indicators constituting it, considered alternatively one at a time: Corruption, Government

effectiveness, Regulatory quality, Rule of law and Voice and accountability. The last rows of all tables report information about sample size and the goodness of fit[9].

Before focusing on the central issue of our exercise, i.e. the effect of institutional quality on students' choice, a preliminary scrutiny of the coefficients of control variables allows us to verify whether and how they are actually relevant to the option on the field of study. The coefficients of regressors representing individual characteristics of students, i.e. gender and parents' education and occupations always assume negative values and are statistically strongly significant (*p values* always lower than 0.01). This means that female students and those with a better cultural and economic familiar background have a higher probability to choose a Defending field of study, substantially confirming our a priori expectations. Since all these regressors are dummy variables, a significant comparison of the relative size of their effects can be simply made by looking at MEMs. This allows to point out that the impact of gender is large (females have a 15% higher probability to choose a Defending field of study) and far more important than the other two variables. These results are very stable across all different specifications of the Probit model (Tables 5 and 6). When using the Logit model (Tables 7 and 8), estimations are again very stable across specifications, and differences with Probit are appreciable for coefficients values but negligible for MEMs. Finally, looking at Tables 9 to 11 (MNL and multilevel estimates), we notice that the estimates on the effect of individual characteristics of students are definitely consistent with results shown in Tables 5 to 8. In addition, the right panel of Table 9 highlights that the sign of the coefficient of Gender turns out to be positive when the choice alternative to *Defending* is *Other fields of study* rather than

[9] For Probit and Logit models (Tables 5-8) the usual statistics are reported, i.e. the value of the likelihood function and the pseudo R^2. Also, the percent correctly predicted is shown (i.e. the percentage of times that the predicted choice matches the actual one). Other measures of the goodness of fit are available upon request. We run tests on multiple exclusion of variables related to i) students' individual characteristics (gender, family education and family network, migratory status), ii) students' secondary education, i.e. High School grade and High School type; iii) the economic condition of students' province of origin (per-capita GDP and industrial intensity) by using both a Wald and Likelihood Ratio test, always rejecting the hypothesis of no effect on students' choice. Concerning the MNL model, (Table 9) we tested the validity of the Independence of Irrelevant Alternatives (IIA) assumption (Cheng and Long, 2006; Kropko 2010) by using a Hausman test, which always yielded negative values. According to Hausman and Mc Fadden (1984), a negative value for the Hausman test can be taken as support for the null hypothesis of IIA validity. The Small-Hsiao test also confirmed the validity of IIA. The likelihood ratio (LR) test run in Tables 10 and 11 shows the significance of regional effects thus confirming the suitability of multilevel modeling.

Productive, showing that being a female student involves a sort of hierarchy in preferences: Literature, Language, Teaching, Psychology first, then *Defending* and finally *Productive*.

The students' high school background also seem to exert an important and statistically significant effect on the students' choice. Results in Tables 5 to 8 show that students with grades not lower than 90/100 have a probability about 10% higher than others to choose a *Productive* rather *Defending* field of study. The impact of having a Liceo (science and literature school) rather than a technical or vocational secondary school degree on the probability to select a *Productive* major is still positive and significant but much lower. Again, estimates are very stable across different specifications and consistent across different estimation models. Like in the case of gender, the comparison between Defending and Other fields of studies made in panel 2 of Table 10 shows a change in the signs of coefficients for both grade and type of secondary school. This seems to indicate that Liceo high-grade students have a consistent order of preferences in the choice of university majors: first *Productive*, then *Defending* and finally *Others*.

Moving to macroeconomic variables, we now look at the impact of local[10] per-capita income, industrial intensity and institutional quality on students' preferences for majors[11]. Concerning provincial per capita GDP, coefficient signs come out to be always positive and significant, confirming a priori expectations about the positive impact of this regressor on the probability of choosing a *Productive* occupation. MEM values are stable across estimation methods but quite variable across specifications. In particular, the quantitative impact of GDP reduces when including among regressors IQI alone (specification 4) and especially IQI and industrial intensity

[10] In particular, we prefer to choose provincial (rather regional) per capita income and industrial intensity because we believe that the provincial level is the most suitable to represent the specific economic condition of families and perceptions on chances of a future occupation in a *Productive* career. On the other hand, we use regional data on the quality of faculties for some provinces do not have universities, and on institutions because it seems likely that expected returns to each career be determined by institutional quality at a wider layer than the provincial one.

[11] The variable representing faculty quality (REEF) is found to be insignificant in most specifications, a relevant exception being multilevel estimates shown in Tables 10 and 11. This result may be due to the relatively low propensity of Italian students to attend at courses outside the region of residence, in turn due to a poor development of credit to education plans.

together (specification 5)[12]. Similar results are delivered by MNL estimation displayed in Table 9: in the left panel, income is always positive and significant but once again the value of its coefficient drastically shrinks when IQI and IND are considered among explanatory variables[13].

The results of regressions show that the overall institutional quality index IQI has a significant effect on the choice of fields of study. For all specifications and any estimation method the evidence is that high institutional quality encourages students to select a *Productive* career versus a *Defending* one (in the case of MNL, even *Others fields of study* versus *Defending*). However, from inspection of Tables 5, 7, 9 and 10 (i.e. those where IQI is employed), the relative importance of institutional quality in shaping students' option seems to be definitely lower than the other macroeconomic variables, as its impact on the probability of choosing *Productive* can be calculated around 9% of the effect of provincial per capita GDP and 7% of industrial intensity (specification 5, Tables 5, 7 and 9).

The impact of institutions is however much higher when the institutional index IQI is replaced by its sub-indexes, which in Tables 6 and 8 are employed one by one as regressors. In this case, a striking result emerges. Indeed the overall impact of IQI seems to derive by different effects of different kinds of institutional factors. So, while some components of IQI do not exert a significant impact on the choice of the field of study, we find, in accordance with expectations, that REG, VOICE and above all RULE have a strong and significant positive impact on the probability of the option *Productive*. In particular RULE, which is calculated on the basis of number of crimes, length of trials, judges' productivity underground economy and tax evasion, shows MEM around 4%, i.e. 1.5 times provincial GDP and 0.7 times industrial intensity (see Tables 6 and 8, column 5) . This outcome is not surprising, considering that RULE basically represents law enforcement, i.e. the main determinants of product appropriability and demand for protection. Thus, law and

[12] A 1% change in provincial per capita GDP involves a change in the probability of choosing a *Productive* occupation between 11.3%, in case IND and IQI are kept out, and 2.6%, when IND and the most important component of IQI, i.e. RULE are considered among regressors.

[13] The estimated effect of per capita GDP is instead unstable in the right panel of Table 9, where the sign surprisingly changes moving from specification (3) and (4) to (5).

sociopolitical studies draw less interest where institutions are stronger, the demand for legal protection and political intermediation is lower and expected rewards for that kind of careers are poor. This econometric exercise shows that the weight of institutional quality on the choice of fields of study is significant, and particularly substantial when the institutional dimension of law enforcement is considered.

4. Concluding remarks

Students' choices about post-secondary fields of study vary widely across space and time, due to many psychological, social and economic motivations. Regarding these latter, the most important role in steering students' options has been often ascribed to expected returns from different occupations. This paper emphasizes in particular the link between local institutional quality, the reward structure and students' preferences. Based on a sample of 80,996 students graduated in Italy in 2004 and 2007, our econometric investigation, controlling for both individual characteristics (gender, residence, family background, high school track) and geographical variables (per capita GDP, industrial specialization), finds that in the choice of the field of study institutional quality definitely matters.

In particular, the overall impact of institutional quality turns out to derive by different effects of different kinds of institutional factors. So, while some components of the overall index do not exert a significant impact on the choice of the field of study, we find that especially the component RULE has a strong and significant positive impact on the probability to enroll in a Science, Chemistry, Pharmacology, Geology, Biology, Engineering, Architecture, Economics and Statistics tertiary course of study (i.e. in the group we name *Productive* studies). This outcome is not surprising, considering that RULE basically represents law enforcement, i.e. the main determinants of product appropriability and demand for protection. Thus, law and sociopolitical studies draw less interest where institutions are stronger, the demand for legal protection and political intermediation is lower and expected rewards for that kind of careers are poor.

Appendix 1

The following framework is inspired to the model proposed by Acemoglu (1995). Here the number of possible occupations is increased to include the group of defending agents, i.e. the ones who choose to protect productive agents from rent-seekers who seek to subtract them a share of the return from productive activity.

We assume that the economy consists of a continuum of identical agents normalized to 1. Each agent can choose to employ her talent in one of the three possible activities: productive, rent-seeking and defending. If involved in the first kind of activity, the agent gains a net return equal to

$$V^P = [1 - \pi q](a + x) - \pi w(p) - c(x) \qquad [A1]$$

where π and p are the shares of rent seekers and defending over the total number of agents; q is the share of resources at stake (i.e. the proportion of production which rent-seekers can seize); w is the wage of defending, assumed to be decreasing in p (i.e. $w'_p < 0$) because of competition among defending agents, with finite $w(0)$; x is investment, $a + x$ its gross revenue and c the production cost, increasing and convex in x, i.e. $c'_x > 0;\ c''_x > 0$.

Without rent-seekers (i.e. $\pi = 0$), the productive agent would get a net revenue $(a + x - c)$; the expected cost from rent-seeking is the sum of gross revenue at risk of expropriation and the wage to pay to defender for protection $q(a + x) + w$, times the probability π that the producer has to deal with a rent-seeker.

The producer determines how much to invest in order to maximize [A1], i.e. the optimal amount of investment such that $c'_x(x) = \{1 - \pi q\}$, which implies that optimal investment $x(\pi, q)$ be everywhere decreasing in π and q.

The net return to being a defending agent is

$$V^D = \pi(1 - \pi - p)w(p) \qquad [A2]$$

i.e. her unit wage times the probability of being employed, which in turn depends on the shares of rent-seekers π and producers $(1 - \pi - p)$.

Finally, the expected return to a rent seeker is

$$V^R = (1 - \pi - p)r(\pi, q) \qquad [A3]$$

i.e. the payoff to rent seekers equals the likelihood of obtaining bribes which in turn is positively connected to the probability of dealing with producers $(1 - \pi - p)$ times the unit reward $r(\pi, q)$, assumed to be decreasing in π because of competition among rent seekers, and increasing in the share of resources at stake, i.e. $r'_\pi < 0;\ r'_q > 0$.

For a given level of q, equilibra with agents choosing all kinds of occupation are characterized by couples (π, p) solving [A1] - [A3], plus the condition of equal expected return to any occupation, i.e.

$$V^P = V^D = V^R = V. \qquad [A4]$$

A possible equilibrium is the one described in Figure 1, where returns to productive, defending and rent-seeking agents are depicted as a function of p and π. To draw Figure 1, we substitute for optimal investment in $x(\pi, q)$ in [A1] and calculate:

$$\frac{\partial V^P}{\partial \pi} = -[a + x(\pi, q)]q - w(p) < 0;$$

$$\frac{\partial^2 V^P}{\partial \pi^2} = -x'_\pi(\pi, q)q > 0;$$

$$\frac{\partial V^D}{\partial \pi} = (1 - 2\pi - p)w(p) > 0$$

and

$$\frac{\partial V^R}{\partial \pi} = -r(\pi, q) + (1 - \pi - p)r'_\pi(\pi, q) < 0.$$

Assuming that $\pi \leq \frac{1-p}{2}$, i.e. the productive agents are not less numerous than rent-seekers, so that $\frac{\partial V^D}{\partial \pi}$ is positive, we can draw the graph of panel (a). A simple argument shows that in this case intersection at point A is a stable equilibrium: in fact, for $\pi > \pi^*$, $V^D > V^P > V^R$, so that rent seeking is comparatively less profitable than the alternative occupations; thus, π decreases. Conversely, if $\pi < \pi^*$, an interval of values of π exists for which $V^R > V^P > V^D$; therefore rent seeking is the most profitable activity and π goes up. Notice that in Figure 1 it is also assumed, like in Acemoglu (1995) that $[a + x(0, q)] - c[x(0, q)] > r(0, q)$. This implies that for π small enough, V^P exceeds returns to any alternative career and an all-producers equilibrium takes place.

To draw the graph of panel (b), we calculate

$$\frac{\partial V^P}{\partial p} = -\pi w'_p(p) > 0;$$

$$\frac{\partial V^D}{\partial p} = -\pi w(p) + \pi(1 - \pi - p)w'_p(p) < 0$$

and

$$\frac{\partial V^R}{\partial p} = -r(\pi, q) < 0$$

and assume that $\pi(1 - \pi)w(0) > (1 - \pi)r(\pi, q) > [1 - \pi q](a + x) - \pi w(0) - c(x) > 0$. In this case, the unique intersection at point A' is a stable equilibrium: for $p < p^*$, being a defending agent

is the most profitable choice and p goes up; for $p > p^*$, the return to lawyers is less than the others; people are discouraged to choose that career and p decreases.

Figure 2 illustrates the effects of an exogenous change in q starting from the equilibrium of Figure 1. As argued in the main text, institutional quality and the amount of production at stake, i.e. the share of resources that rent-seekers can seize, are closely (inversely) related. So, an improvement of institutional quality may be represented as a reduction of q, yielding:

$$\frac{\partial V^P}{\partial q} = -\pi[a + x(\pi, q)] < 0;$$

$$\frac{\partial V^R}{\partial q} = (1 - \pi - p)r_q' > 0.$$

Graphically, the effects of decreasing q are illustrated by shifts in V^R (downward, because rent seekers get less from each bribe) and V^P (upward, because producers lose less). As a result, a new stable equilibrium in BB' is reached with lower shares of rent seekers and defending agents and a higher share of productive agents. Notice that even if V^D does not depend on q, in response to an improvement in institutional quality, it shifts as well toward the new equilibrium BB', because of changes occurring in p and in π (precisely upward in panel (a) and downward in panel (b)). However, for the sake of clarity, these modifications are not shown in Figure 2.

Appendix 2

Data sources.

For variables 1-8 of Table 1, sources are ISTAT (2009) and (2013); for variable 9, ANVUR (2013); for variable 10, ISTAT (2011); for variable 11, Istituto Tagliacarne (2004); for variables 12-17, Nifo and Vecchione (2014). Concerning in particular variables 13-17, CORR is composed by elementary indicators on: crimes against PA (ratio crimes against the public administration over number of public servants; ISTAT, 2004); the Golden-Picci corruption index (Golden and Picci, 2005); special commissioners (number of overruled municipalities on total municipalities; Ministero dell'Interno, 2006). GOV is made up of: endowment of social facilities (education, healthcare and leisure facilities; Istituto Tagliacarne, 2001); endowment of economic facilities (roads, railroads, ports, airports, energy, ICT, banking; Istituto Tagliacarne, 2004); public health care budget deficit (per capita 1997-2004; elaboration on Ministero dell'Economia, 2004); waste sorting (share of waste sorting on total waste; Istituto Tagliacarne, 2004); urban environment index (including 25 indexes on: air quality, water quality, purification plants, waste management, public transportation, energy consumption, public parks, eco-management, etc. Legambiente, 2004). REG

comprises elementary indexes on economy openness (ratio import plus export on local GDP; Istituto Tagliacarne, 2004); employed by local governments (number of public servants over resident population; ISTAT, 2003); business density (number of firms for 100 residents; Istituto Tagliacarne, 2004); business start-ups/mortality (registration/mortality rate; Istituto Tagliacarne, 2004); business environment (including 39 indexes on: entrepreneurship, job market, tax system, market competition, banking, bureaucracy, public services to firms, firms' cooperation; Confartigianato, 2009). RULE is composed by the number of crimes against property (ratio crimes against property over resident population; ISTAT, 2003) and total reported crimes (ratio number of reported crimes over resident population; ISTAT, 2003); length of trials (average length of judicial process; CRENOS, 2001); judges' productivity (number of completed trials for magistrate in regional courts; Ministero della Giustizia, 2009); underground economy (ISTAT, 2003); tax evasion (Agenzia delle entrate, 2006). VOICE is made up of: social cooperatives (per 100,000 residents; ISTAT, 2005); associations (per 100,000 residents; ISTAT, 2006); election turn-out (general election in 2001; Ministero dell'Interno, 2013); number of published books (in absolute value; ISTAT, 2007); purchases of books (number of purchased books over resident population; Il Sole24Ore, 2004).

References

Acemoglu, D. (1995). "Reward structures and the allocation of talent", *European Economic Review*, 39(1), pp. 17–33.

Acemoglu, D., and Pischke, S. (2001). "Changes in the wage structure, family income and children's education", *European Economic Review*, 45(4-6), pp. 890–904.

Agenzia delle entrate (2006). *Analisi dell'evasione fondata su dati IRAP. Anni 1998-2002*, Roma.

Akerlof, G.A., and Kranton, R.E. (2002). "Identity and schooling: Some lessons for the economics of education", *Journal of Economic Literature*, 40(4), pp. 1167-1201.

Altonji, J., Blom, E., and Meghir, C. (2011). "Heterogeneity in human capital investments: High school curriculum, college major, and careers", *Annual Review of Economics*, 4(1), PP. 185–223.

ANVUR (2013). *Rapporto finale. Rapporti di area 9 e area 12*. Downloadable from: http://www.anvur.org/rapporto/

Arruñada, B. (2007), "Pitfalls to Avoid when Measuring Institutions: Is 'Doing Business' Damaging Business?", *Journal of Comparative Economics,* 35, pp. 729-747.

Averett, S., and Burton, M. (1996). College attendance and college wage premium: Differences by gender. *Economics of Education Review*,15(1), pp. 37–49.

Banerjee, A.V., and Newman, A.F. (1993). "Occupational choice and the process of development", *Journal of Political Economy*, 101(2), pp. 274–298.

Becker, G.S. (1964). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. New York: Columbia University Press.

Beffy, M., Fougère, D., and Maurel, A. (2012). "Choosing the field of study in postsecondary education: Do expected earnings matter?", *The Review of Economics and Statsitics*, 94(1), pp. 334-347.

Berger, M.C. (1988). "Predicted Future Earnings and Choice of College Major", *Industrial and Labor Relations Review*, 41, pp. 418-429.

Boeri, T., Patacchini, E., and Peri, G. (2015). *Unexplored Dimensions of Discrimination*, Oxford University Press.

Boudarbat, B. (2008). "Field of study choice by community college students in Canada", *Economics of Education Review*, 27, pp. 79–93.

Boudarbat, B., and Montmarquette, C. (2009). "Choice of fields of study of university Canadian graduates: The role of gender and their parents' education," *Education Economics*, 17(2), pp. 185-213.

Broecke, S. (2015). "University rankings: do they matter in the UK?", *Education Economics*, 23(2), pp. 137-161.

Buser, T., Niederle, M., and Oosterbeek, H. (2014). "Gender, competitiveness, and career choices", *Quarterly Journal of Economics*, 129(3), pp. 1409-1447.

Cameron, C.A., and Trivedi, P.K. (2010). "Microeconometrics Using Stata", Stata Press.

Carneiro, P., and Heckman, J. (2002). "The Evidence on Credit Constraints in Post-Secondary Schooling", *Economic Journal*, 112(482), pp. 705-734.

Cersosimo, D., and Viesti, G. (2013). "High Tech and Mezzogiorno: Industrial development and public policies", *QA Rivista dell'Associazione Rossi-Doria*, 1, pp. 7-45.

Charles, M., and Grusky, D.B. (2004), *Occupational Ghettos: The Worldwide Segregation of Women and Men*, Stanford University Press.

Checchi, D., and Flabbi, L. (2013). "Intergenerational mobility and schooling decisions in Germany and Italy", *Rivista di Politica Economica*, VIII-IX, pp.1-52.

Cheng, S., and Scott Long, J. (2006) "Testing for IIA in the Multinomial Logit Model." Sociological Methods & Research, 35, pp. 583-600.

Clarke, M. (2007). "The impact of higher education rankings on student access, choice, and opportunity", *Higher Education in Europe*, 32(1), pp. 59-70.

Confartigianato (2009), L'indice Confartigianato – Qualità della vita dell'impresa, Roma.

Corak, M. (2006). "Do Poor Children Become Poor Adults? Lessons for public policy from a Cross Country Comparison of Generational Earnings Mobility", *Research on Economic Inequality*, 13, pp. 143-188.

CRENOS (2001), Data-base on crime and deterrence in the Italian regions (1970-1999), Cagliari.

Cumming, D., and Johan S. (2006). "Is it the Law or the Lawyers? Investment Covenants around the World", *European Financial Management,* 12, pp. 535-574.

Dezalay, Y., and Garth, B. (1997). "Law. Lawyers, and Social Capital: Rule of Law Versus Relational Capitalism" , *Social & Legal Studies,* 6, pp. 109-141.

Dustmann, C. (2004). "Parental Background, Secondary School Track Choice, and Wages", *Oxford Economic Papers*, 56, pp. 209-230.

Esping-Andersen G. (2004). "Unequal Opportunities and the Mechanisms of Social Inheritance", in Corak M. (ed.), Generational Income Mobility in North America and Europe, Cambridge University Press.

European Commission (2009), Women in science and technology. Creating sustainable careers, Report EUR 23740 EN, downloaded from: http://ec.europa.eu/research/science-society/document_library/pdf_06/wist2_sustainable-careers-report_en.pdf

Fiorito, J., and Dauffenbach, R. C. (1982). "Market and non market influences on curriculum choice by college students", *Industrial and Labor Relations Review*, 36(1), pp. 88–101.

Golden M.A. and Picci L. (2005), Proposal for a new measure of corruption, illustrated with Italian data, *Economics and Politics*, 17, pp. 37-75.

Goldin C. (2006) "The 'Quiet Revolution' that Transformed Women's Employment, Education and Family", *American Economic Review*, 96 (2), pp. 1-21.

Goldstein, H. (2003). "Multilevel Statistical Models", Arnold, London.

Goodman, J., Hurwitz, M., Smith, J., and Fox J. (2015), "The relationship between siblings' college choices: Evidence from one million SAT-taking families", *Economics of Education Review* 48, pp. 75–85.

Hall, R.E., and Jones, C.I. (1999). "Why do some countries produce so much more output per worker than others?", *Quarterly Journal of Economics*, 114(1), pp. 83–116.

Hausman, J.A., and McFadden, D. (1984). "Specification tests for the multinomial logit model", *Econometrica*, 52, pp. 1219-1240.

Haveman, R., and Wolfe, B. (1995). "The Determinants of Children's Attaiments: A Review of Methods and Findings", *Journal of Economic Literature*, 33(4), pp. 1829- 1878.

Hertz, T., Jayasundera, T., Piraino, P., Selcuk, S., Smith, N., and Verashchagina A. (2007). "The Inheritance of Educational Inequality: International Comparisons and FiftyYear Trends", *The B.E. Journal of Economic Analysis & Policy*, 7(2).

Hilmer, M. (1998). "Post-Secondary Fees and the Decision to Attend a University or a Community College", *Journal of Public Economics*, 67, pp. 329–348.

House of Commons. Science and Technology Committee (2014). *Women in scientific careers*, Sixth Report of Session 2013–14, January, London. Downloaded from: http://www.publications.parliament.uk/pa/cm201314/cmselect/cmsctech/701/701.pdf

Hox, J.J. (2010). "Multilevel Analysis. Techniques and Applications", Routledge, New York.

Humlum, M.K., Kleinjans, K.J., and Nielsen, H.S. (2012). "An economic analysis of identity and career choice", *Economic Inquiry*, 50(1), 39–61.

Il Sole24Ore (2004), Rapporto 2004 sulla qualità della vita, Roma.

ISTAT (2003), Indicatori statistici sulle amministrazioni centrali e locali. Downloaded from: http://dati.statistiche-pa.it/

ISTAT (2003), Indicatori territoriali per le politiche di sviluppo, Roma.

ISTAT (2003), Le misure dell'economia sommersa secondo le statistiche ufficiali, Roma.

ISTAT (2004), Indicatori territoriali per le politiche di sviluppo, Roma.

ISTAT (2005), Le organizzazioni di volontariato in Italia, Roma.

ISTAT (2006), Le cooperative sociali in Italia, Roma.

ISTAT (2007), La produzione libraria, Roma.

ISTAT (2009) Indagine ISTAT sull'inserimento professionale dei laureati dell'anno 2004, wave VII, Roma.

ISTAT (2010) Conti economici regionali 1995-2009, Roma.

ISTAT (2011) Sistemi di Indicatori Territoriali 1998-2010, www.istat.it , Roma.

ISTAT (2013) Indagine ISTAT sull'inserimento professionale dei laureati dell'anno 2007, wave VIII, Roma.

Istituto Tagliacarne (2004), Atlante di competitività delle province italiane (2003-2004), Roma.

Jacobsen, J.P. (2007). *The Economics of Gender*, 3rd. ed., Malden, MA, Blackwell.

Keefer, P., and Knack, S. (2007), "Boondoggles, Rent-Seeking, and Political Checks and Balances: Public Investment under Unaccountable Governments", *Review of Economics and Statistics*, 89 (3), pp. 566-572.

Kirkebøen, L., Leuven, E., and Mogstad, M. (2014). "Field of study, earnings, and self-selection", *NBER Working Papers*, 20816, December.

Kreft, I.G.G., and de DeLeeuw, J. (1998). *Introducing Multilevel Modeling*, Sage, London.

Kropko, J. (2010). A comparison of three discrete choice estimators. Unpublished paper. Available from: http://www.unc.edu/kropko/paper1.pdf

Krueger, A. (1974), "The political economy of the rent-seeking society", *American Economic Review*, 64 (3), pp. 291-303.

Legambiente (2004), Ecosistema Urbano 2004, Roma.

Levchenko, A. (2007). "Institutional Quality and International Trade," *Review of Economic Studies,* 74, pp. 791–819.

Long, B. (2004). "How have college applications changed over time? An application of the conditional logistic choice model", *Journal of Econometrics,* 121, pp. 271-296.

Luca, M., and Smith, J. (2013). "Salience in quality disclosure: Evidence from the U.S. news college rankings", *Journal of Economics and Management Strategy,* 22(1), pp. 58-77.

Luke, A. (2004). Multilevel Modeling. Sage, London.

Magee, S.P. (2010). "The optimal number of lawyers and a radical proposal for legal change", paper presented at the Conference *An American Illness: Essays on the Rule of Law*, School of Law, George Mason University, December 3-4, Downloadable at: http://buckleysmix.com/wp-content/uploads/2010/10/Magee.pdf

Magee, S.P., Brock, W.A., and Young, L. (1989). *Black Hole Tariffs and Endogenous Policy Theory: Political Economy in General Equilibrium*, New York: Cambridge University Press.

Mehlum, H., Moene, K., and Torvik, R. (2003). "Predator or prey? Parasitic enterprises in economic development", *European Economic Review*, 47(2), pp. 275–294.

Ministero della Giustizia (2009), Graduatoria rispetto agli esauriti per magistrato presente (2004-2008), Roma.

Ministero dell'Economia (2004), Relazione generale sulla situazione economica del Paese (1997-2004), Roma.

Ministero dell'Interno (2006), Relazione sull'attività svolta dalla gestione straordinaria dei Comuni commissariati (1991-2005), Roma.

Ministero dell'Interno (2013), Archivio storico delle elezioni, Downloaded from: http://elezionistorico.interno.it/

Montmarquette, C., Cannings, K., and Mahseredjian, S. (2002). "How do Young People Choose College Majors?", *Economics of Education Review*, 21(6), pp. 543-556.

Muller, R.E., and Rockerbie, D. (2005). "Determining demand for university education in Ontario by type of student", *Economics of Education Review*, 24, pp. 469-483.

Murphy, K., Shleifer, A., and Vishny, R. (1991). "The allocation of talent: Implications for growth", *Quarterly Journal of Economics,* 106, pp. 503–530.

Mushtaq, H.K., and Jomo, K.S. (2000). *Rents, rent seeking and economic development. Theory and evidence in Asia*, Cambridge University Press.

Natkhov,T., and Polishchuk, L. (2013). "Institutions and the allocation of talent", WP 15/EC/2012, National Research University, Higher School of Economics, Moscow.

Nguyen, A.N., and Taylor, J. (2003). "Post-high school choices: New evidence from a multinomial logit model", *Journal of Population Economics*, 16, pp. 287–306.

Nifo, A., and Vecchione, G. (2014). "Do institutions play a role in skilled migration? The case of Italy", *Regional Studies*, vol. 48(10), pp. 1628-1649.

North, D.C. (1992). "The new institutional economics and development", Washington University, St. Louis. Downloadable from: http://128.118.178.162/eps/eh/papers/9309/9309002.pdf

Nunn, N. (2007). "Relationship-Specificity, Incomplete Contracts, and the Pattern of Trade," *Quarterly Journal of Economics*, 122, pp. 569–600.

Pecorino, P. (1992). "Rent seeking and growth: The case of growth through human capital accumulation", *Canadian Journal of Economics*, 25(4), pp. 944–956.

Pritchett, L. (2006). "Does learning to add up add up? The returns to schooling in aggregate data," in *Handbook of the Economics of Education*, Hanushek, E. and Welch, F. (eds.), Elsevier.

Rosenbloom, J.L., Ash, R.A., Dupont, B., and Coder, L.A. (2008). "Why are there so few women in information technology? Assessing the role of personality in career choices", *Journal of Economic Psychology*, 29(4), pp. 543–554.

Siltanen, J., Jarman, J., and Blackburn, R.M. (1995). *Gender Inequality in the Labour Market, Occupational Concentration and Segregation*. Geneva: International Labour Office.

Tollison, R., (1982). "Rent Seeking: A Survey," *Kyklos,* 35, pp. 575-602.

UK Government (2011). *Women on boards*, London, downloadable from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/31480/11-745-women-on-boards.pdf

UNESCO Institute for Statistics (2014). *Distribution of tertiary graduates by field of study*, downloadable from: http://data.uis.unesco.org/Index.aspx?queryid=163

Wang, X. (2013). "Why students choose STEM majors: Motivation, high school learning, and postsecondary context of support", *American Educational Research Journal*, 50, pp. 1081-1121.

Weichselbaumer, D., and Winter-Ebmer, R. (2005). "A Meta-Analysis of the international gender wage gap", *Journal of Economic Surveys*, 19(3), pp.479-511.

Wiswall, M., and Zafar, B. (2015). "Determinants of College Major Choice: Identification using an Information Experiment", *The Review of Economic Studies*, 82(2), pp. 791-824.

Xie, Y., and Shauman, K. (2003). *Women in Science: Career Processes and Outcomes*, Harvard University Press.

Tables and Figures

Table 1**.** Dependent and explanatory variables of the econometric investigation.

Variable n	name	Description	Average	Min	Max
Dependent variables					
1.	y_1 (Probit and Logit)	Dichotomous dummy. 0=Law; 1=Science.	0.704	0	1
2.	y_2 (MNL)	Trichotomous dummy. 0=Law; 1=Science; 2=Other fields of study.	1.526	0	2
Individual characteristics					
3.	Gender	Dichotomous dummy. 0=Male; 1=Female.	0.541	0	1
4.	Par_Edu	Dichotomous dummy. 0=Neither parent graduated ;1= One or both parents graduated.	0.252	0	1
5.	Par_occ	Dichotomous dummy. 0=Neither parent is a professional or entrepreneur; 1=One or both parents are professionals or entrepreneurs.	0.401	0	1
Secondary education					
6.	HS_grade	Dichotomous dummy. 0=final Grade Point Average at High School $< 90/100$; 1=final Grade Point Average at High School $\geq 90/100$	0.411	0	1
7.	HS_type	Dichotomous dummy. 0= other High Schools; 1= "Liceo" High Schools.	0.580	0	1
Tertiary education					
8.	REEF	Relative evaluation of Engineering Faculty. Regional average.	1.024	0.211	2.022
Economic condition					
9.	GDP	Provincial per capita GDP. Yearly averages 2001-2004, current value Euros.	19,149	9,947	32,370
10.	IND	Provincial ratio industrial employees over population	0.091	0.043	0.308
Institutional quality					
11.	IQI	Regional institutions index	0.572	0.141	0.900
12.	CORR	Regional (inverse) corruption index, log	0.448	0.305	0.969
13.	GOV	Regional policy effectiveness index	0.245	0.082	0.514
14.	REG	Regional regulatory quality index	0.386	0.199	0.850
15.	RULE	Regional rule of law index	0.643	0.192	0.863
16.	VOICE	Regional civil society index	0.336	0.173	0.695

Data sources: See Appendix 2

Table 2. Graduates for field of study and group in Italy, number and % share.

Field of study	**2004**	**%**	**2007**	**%**	**Group**	**2004**	%	**2007**	%
Science	6555	2.52	7800	2.60					
Chemistry and pharmacology	6834	2.63	6863	2.29					
Geo-Biological	10904	4.19	13787	4.59	**Productive**	108735	41.81	120651	40.17
Engineering	32328	12.43	35766	11.91					
Architecture	11708	4.50	15065	5.02					
Economics and statistics	40406	15.54	41370	13.77					
Law	28406	10.92	25604	8.53	**Defending**	57193	21.99	65724	21.88
Political-Social	28787	11.07	40120	13.36					
Agriculture	5416	2.08	5662	1.89					
Literature	21811	8.39	25728	8.57					
Language	12699	4.88	15770	5.25					
Teaching	12745	4.90	15681	5.22	**Others**	65939	25.35	80909	26.94
Psychology	10119	3.89	13416	4.47					
Physical education	2762	1.06	3797	1.26					
Defence-Security	387	0.15	855	0.28					
Medical studies	28203	10.84	33054	11.01	**Medical studies**	28203	10.84	33054	11.01
Total	**260070**	**100.00**	**300338**	**100.00**	**Total**	**260070**	**100.00**	**300338**	**100.00**

Data source: ISTAT (2009, 2013)

Table 3. Graduates for field of study and macro-region in Italy, number and % share.

Field of study	NW	NE	Centre	South	Abroad	NR	NW	NE	Centre	South	Abroad	NR
	Number						% share					
Science	707	630	493	551	55	25	3.37	3.96	2.78	2.24	5.40	3.65
Chemistry and pharmacology	746	534	503	799	61	25	3.55	3.35	2.83	3.24	5.99	3.65
Geo-Biological	850	530	687	1028	54	23	4.05	3.33	3.87	4.17	5.30	3.36
Engineering	2521	1919	1993	2565	155	67	12.01	12.05	11.23	10.41	15.21	9.80
Architecture	1317	910	745	1050	64	36	6.28	5.71	4.20	4.26	6.28	5.26
Economics and statistics	2928	2177	2371	3144	158	115	13.95	13.67	13.36	12.76	15.51	16.81
Law	1560	1218	1390	2724	44	36	7.43	7.65	7.83	11.06	4.32	5.26
Political-Social	1641	1122	1566	2127	114	94	7.82	7.04	8.82	8.64	11.19	13.74
Agriculture	367	393	340	584	35	9	1.75	2.47	1.92	2.37	3.43	1.32
Literature	834	727	871	1114	65	37	3.97	4.56	4.91	4.52	6.38	5.41
Language	665	542	484	632	96	72	3.17	3.40	2.73	2.57	9.42	10.53
Teaching	648	663	471	842	15	19	3.09	4.16	2.65	3.42	1.47	2.78
Psychology	487	378	407	541	15	21	2.32	2.37	2.29	2.20	1.47	3.07
Physical education	759	438	415	709	5	4	3.62	2.75	2.34	2.88	0.49	0.58
Medical studies	4903	3696	4942	6117	82	100	23.36	23.20	27.85	24.83	8.05	14.62
No Response	52	52	70	104	1	1	0.25	0.33	0.39	0.42	0.10	0.15
Total	**20985**	**15929**	**17748**	**24631**	**1019**	**684**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**

Data source: ISTAT (2009, 2013)

Table 4. Shares of graduates for group, macro-region and wave

Wave	I Wave, graduated in 2004				II Wave, graduated in 2007			
Group	North-W	North-E	Centre	South	North-W	North-E	Centre	South
Productive	55.73	52.87	51.67	47.73	57.62	57.96	55.17	50.93
Defending	19.51	18.62	22.34	25.46	20.63	19.98	24.25	26.92
Others	24.76	28.52	25.99	26.81	21.76	22.06	20.59	22.15
Total	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**

Data source: ISTAT (2009, 2013)

Table 5. Probit Model. Estimated coefficients and marginal effects (0: *Defending*; 1: *Productive*)

	(1)		(2)		(3)		(4)		(5)	
	Coeff	MEM	Coeff	MEM	Coeff	MEM	Coeff	MEM	Coeff	MEM
Gender	-0.472***	-0.158***	-0.464***	-0.156***	-0.459***	-0.154***	-0.459***	-0.153***	-0.461***	-0.154***
	(0.0129)	(0.00414)	(0.0133)	(0.00429)	(0.0133)	(0.00431)	(0.0133)	(0.00432)	(0.0134)	(0.00431)
Par_Edu	-0.0952***	-0.0320***	-0.0903***	-0.0303***	-0.0866***	-0.0291***	-0.0867***	-0.0297***	-0.0868***	-0.0295***
	(0.0171)	(0.00572)	(0.0177)	(0.00591)	(0.0176)	(0.00592)	(0.0177)	(0.00592)	(0.0177)	(0.00591)
Par_Occ	-0.0559***	-0.0188***	-0.0593***	-0.0199***	-0.0719***	-0.0241***	-0.0725***	-0.0243***	-0.0685***	-0.0229***
	(0.0149)	(0.00500)	(0.0154)	(0.00516)	(0.0154)	(0.00518)	(0.0154)	(0.00517)	(0.0154)	(0.00517)
HS_grade	0.2818***	0.0946***	0.294***	0.0989***	0.297***	0.0996***	0.298***	0.1001***	0.299***	0.01001***
	(0.0131)	(0.00432)	(0.0136)	(0.00447)	(0.0136)	(0.00448)	(0.0135)	(0.00447)	(0.0136)	(0.00447)
HS_type	0.0307**	0.0103**	0.0327**	0.0111**	0.0343**	0.0115**	0.0347**	0.0114**	0.0404***	0.0135***
	(0.0136)	(0.00457)	(0.0141)	(0.00473)	(0.0144)	(0.00474)	(0.0141)	(0.00473)	(0.0142)	(0.00474)
lg_REEF			-0.0409	-0.0137	0.0252	0.00846	0.0197	0.0066	0.0219	0.00883
			(0.0273)	(0.00912)	(0.0278)	(0.00934)	(0.0278)	(0.00933)	(0.0279)	(0.00934)
lg_GDP					0.339***	0.113***	0.282***	0.0945***	0.1331***	0.0444***
					(0.0241)	(0.00805)	(0.0256)	(0.0085)	(0.0314)	(0.0139)
lg_IND									0.1805***	0.0603***
									(0.0231)	(0.0073)
lg_IQI							0.0131***	0.0044***	0.0123***	0.0041***
							(0.0244)	(0.0006)	(0.0019)	(0.0006)
Constant	0.672***		0.661***		-2.685***		-2.108		0.2083	
	(0.0133)		(0.0137)		(0.242)		(0.417)		(0.457)	
N	44222		41410		41087		41087		41087	
Pseudo-R^2	0.0312		0.0310		0.0346		0.0364		0.0371	
Log likelihood	-26141.974		-24528.25		-24256.68		-24211.632		-24194.233	
Correctly predicted	70.08%		69.95%		69.84%		69.94%		69.91%	

Note: Standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6. Probit Model. Estimated coefficients and marginal effects (0: *Defending*; 1: *Productive*)

	(1)		(2)		(3)		(4)		(5)	
	Coeff	MEM	Coeff	MEM	Coeff	MEM	Coeff	MEM	Coeff	MEM
Gender	-0.461***	-0.154***	-0.461***	-0.154***	-0.461***	-0.154***	-0.462***	-0.154***	-0.461***	-0.154***
	(0.0135)	(0.00432)	(0.0135)	(0.00431)	(0.0134)	(0.00432)	(0.0135)	(0.00431)	(0.0135)	(0.00432)
Par_Edu	-0.0867***	-0.0291***	-0.0865***	-0.0289***	-0.0866***	-0.0289***	-0.0887***	-0.0296***	-0.0865***	-0.0289***
	(0.0178)	(0.00592)	(0.0178)	(0.00592)	(0.0177)	(0.00592)	(0.0178)	(0.00592)	(0.0178)	(0.00592)
Par_Occ	-0.0671***	-0.0226***	-0.0672***	-0.0227***	-0.0676***	-0.0226***	-0.0661***	-0.0221***	-0.0677***	-0.0226***
	(0.0156)	(0.00517)	(0.0156)	(0.00517)	(0.0154)	(0.00517)	(0.0156)	(0.00517)	(0.0156)	(0.00517)
HS_grade	0.297***	0.0995***	0.297***	0.0996***	0.297***	0.0996***	0.297***	0.0996***	0.297***	0.0995***
	(0.0137)	(0.00447)	(0.0137)	(0.00447)	(0.0135)	(0.00447)	(0.0137)	(0.00447)	(0.0137)	(0.00447)
HS_type	0.0403***	0.0135***	0.0404***	0.0135***	0.0403***	0.0135***	0.0395***	0.0132***	0.0401***	0.0134***
	(0.0142)	(0.00474)	(0.0142)	(0.00474)	(0.0142)	(0.00474)	(0.0142)	(0.00474)	(0.0142)	(0.00474)
lg_REEF	0.0278	0.00932	0.0244	0.00817	0.0266	0.00891	0.00834	0.000279	0.0265	0.00887
	(0.0282)	(0.00937)	(0.0282)	(0.00941)	(0.0279)	(0.00935)	(0.0282)	(0.00939)	(0.0279)	(0.00935)
lg_GDP	0.179***	0.0601***	0.185***	0.0624***	0.174***	0.0584***	0.0806**	0.0269**	0.161***	0.0542***
	(0.0308)	(0.0103)	(0.0311)	(0.0103)	(0.0307)	(0.0102)	(0.0345)	(0.0114)	(0.0323)	(0.0108)
lg_IND	0.186***	0.0624***	0.189***	0.0633***	0.181***	0.0603***	0.169***	0.0567***	0.189***	0.0632***
	(0.0222)	(0.00743)	(0.0223)	(0.00743)	(0.0226)	(0.00752)	(0.0224)	(0.00744)	(0.0223)	(0.00741)
lg_CORR	0.00061	0.000205								
	(0.00180)	(0.000601)								
lg_GOV			-0.00364	-0.00122						
			(0.00402)	(0.00134)						
lg_REG					0.00573*	0.00192*				
					(0.0034)	(0.00116)				
lg_RULE							0.121***	0.0401***		
							(0.0185)	(0.00619)		
lg_VOICE									0.0103*	0.00347*
									(0.00592)	(0.00198)
Constant	-0.2338		-0.2835		-0.209		0.715*		-0.038	
	(0.375)		(0.375)		(0.378)		(0.402)		(0.392)	
N	41087		41087		41087		41073		41087	
Pseudo-R^2	0.0360		0.0360		0.0361		0.0369		0.0361	
Log likelihood	-24220.95		-24220.592		-24219.666		-24192.404		-24219.482	
Correctly predicted	69.83%		69.82%		69.84%		69.87%		69.83%	

Note: Standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7. Logit Model. Estimated coefficients and marginal effects (0: *Defending*; 1: *Productive*)

	(1)		(2)		(3)		(4)		(5)	
	Coeff	MEM	Coeff	MEM	Coeff	MEM	Coeff	MEM	Coeff	MEM
Gender	-0.782***	-0.157***	-0.769***	-0.155***	-0.761***	-0.153***	-0.761***	-0.153***	-0.763***	-0.153***
	(0.0216)	(0.00413)	(0.0224)	(0.00428)	(0.0223)	(0.00429)	(0.0225)	(0.00429)	(0.0225)	(0.00429)
Par_Edu	-0.155***	-0.0314***	-0.147***	-0.0298***	-0.142***	-0.0286***	-0.145***	-0.0286***	-0.142***	- 0.0286***
	(0.0283)	(0.00571)	(0.0294)	(0.00591)	(0.0295)	(0.00591)	(0.0296)	(0.00591)	(0.0296)	(0.00591)
Par_Occ	-0.0930***	-0.0187***	-0.0986***	-0.0199***	-0.119***	-0.0241***	-0.121***	-0.0226***	-0.114***	-0.0229***
	(0.0248)	(0.00499)	(0.0257)	(0.00515)	(0.0259)	(0.00517)	(0.0257)	(0.00517)	(0.0259)	(0.00517)
HS_grade	0.466***	0.0940***	0.486***	0.0983***	0.491***	0.0989***	0.494***	0.0994***	0.495***	0.0994***
	(0.0219)	(0.00433)	(0.0227)	(0.00448)	(0.0229)	(0.00449)	(0.0227)	(0.00449)	(0.0229)	(0.00449)
HS_type	0.0578**	0.0116**	0.0615***	0.0124***	0.0641***	0.0129***	0.0646***	0.0131***	0.0747***	0.0151***
	(0.0227)	(0.00457)	(0.0235)	(0.00473)	(0.0236)	(0.00474)	(0.0235)	(0.00473)	(0.0237)	(0.00474)
lg_REEF			-0.0642	-0.0129	0.0408	0.0096	0.0386	0.0077	0.0409	0.0082
			(0.0454)	(0.00915)	(0.0468)	(0.00938)	(0.0466)	(0.0093)	(0.0470)	(0.00941)
lg_GDP					0.563***	0.113***	0.4689***	0.0943***	0.2199***	0.0442***
					(0.0407)	(0.00805)	(0.0427)	(0.0085)	(0.0523)	(0.0105)
lg_IND									0.2991***	0.0601***
									(0.0371)	(0.0074)
lg_IQI							0.0213***	0.0042***	0.0201***	0.0045***
							(0.0032)	(0.0006)	(0.0418)	(0.0006)
Constant	1.094***		1.074***		-4.474***		-3.526***		0.3308***	
	(0.0223)		(0.0237)		(0.402)		(0.423)		(0.766)	
N	44222		41410		41087		41087		41087	
Pseudo-R^2	0.0310		0.0308		0.0344		0.0362		0.0369	
Log likelihood	-26146.031		-24532.085		-24261.429		-24216.804		-24199.847	
Correctly predicted	70.08%		69.95%		69.86%		69.94%		69.96%	

Note: Standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8. Logit Model. Estimated coefficients and marginal effects (0*: Defending*; 1: *Productive*)

	(1)		(2)		(3)		(4)		(5)	
	Coeff	MEM	Coeff	MEM	Coeff	MEM	Coeff	MEM	Coeff	MEM
Gender	-0.763***	-0.153***	-0.763***	-0.153***	-0.764***	-0.153***	-0.762***	-0.153***	-0.763***	-0.153***
	(0.0225)	(0.00429)	(0.0225)	(0.00429)	(0.0225)	(0.00429)	(0.0225)	(0.00429)	(0.0225)	(0.00429)
Par_Edu	-0.142***	-0.0286***	-0.142***	-0.0285***	-0.142***	-0.0285***	-0.145***	-0.0295***	-0.142***	-0.0285***
	(0.0296)	(0.00591)	(0.0294)	(0.00591)	(0.0296)	(0.00591)	(0.0296)	(0.00591)	(0.0296)	(0.00591)
Par_Occ	-0.113***	-0.0227***	-0.113***	-0.0227***	-0.112***	-0.0227***	-0.111***	-0.0222***	-0.112***	-0.0227***
	(0.0259)	(0.00517)	(0.0259)	(0.00517)	(0.0259)	(0.00517)	(0.0259)	(0.00517)	(0.0259)	(0.00517)
HS_grade	0.492***	0.0989***	0.492***	0.0991***	0.492***	0.0991***	0.492***	0.0991***	0.492***	0.0991***
	(0.0229)	(0.00449)	(0.0229)	(0.00449)	(0.0229)	(0.00449)	(0.0229)	(0.00449)	(0.0229)	(0.00449)
HS_type	0.0746***	0.0151***	0.0746***	0.0150***	0.0745***	0.0151***	0.0731***	0.0146***	0.0741***	0.0149***
	(0.0237)	(0.00474)	(0.0237)	(0.00474)	(0.0237)	(0.00474)	(0.0237)	(0.00474)	(0.0237)	(0.00476)
lg_REEF	0.0511	0.00102	0.0455	0.00916	0.0491	0.00985	0.0174	0.00349	0.0488	0.00981
	(0.0472)	(0.00944)	(0.0473)	(0.00946)	(0.0470)	(0.00941)	(0.0473)	(0.00945)	(0.0470)	(0.00941)
lg_GDP	0.296***	0.0596***	0.305***	0.0614***	0.288***	0.0581***	0.1313**	0.0263**	0.268***	0.0539***
	(0.0518)	(0.0103)	(0.0518)	(0.0103)	(0.0517)	(0.0102)	(0.0574)	(0.0114)	(0.0542)	(0.0108)
lg_IND	0.308***	0.0621***	0.313***	0.0629***	0.298***	0.0601***	0.281***	0.0565***	0.312***	0.0629***
	(0.0374)	(0.00746)	(0.0375)	(0.00748)	(0.0379)	(0.00757)	(0.0375)	(0.00752)	(0.0373)	(0.00744)
lg_CORR	0.001073	0.00021								
	(0.00295)	(0.000592)								
lg_GOV			-0.00551	-0.00111						
			(0.0065)	(0.00132)						
lg_REG					0.0092*	0.00185*				
					(0.00570)	(0.00114)				
lg_RULE							0.197***	0.0397***		
							(0.0306)	(0.00610)		
lg_VOICE									0.0166*	0.00334*
									(0.00950)	(0.00191)
Constant	-0.396		-0.473		-0.359		1.191*		-0.083	
	(0.625)		(0.632)		(0.630)		(0.676)		(0.658)	
N	41087		41087		41087		41073		41087	
Pseudo-R^2	0.0358		0.0358		0.0359		0.0366		0.0359	
Log likelihood	-24226.686		-24226.399		-24225.459		-24198.004		-24225.24	
Correctly predicted	69.80%		69.81%		69.81%		69.87%		69.82%	

Note: Standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 9. Multinomial Logit Model. Estimated coefficients (0: *Defending*).

	(1)	(2)	(3)	(4)	(5)		(1)	(2)	(3)	(4)	(5)
1: Productive						2: Others					
Gender	-0.780*** (0.0213)	-0.767*** (0.0221)	-0.761*** (0.0223)	-0.761*** (0.0223)	-0.763*** (0.0224)	Gender	0.266*** (0.0250)	0.286*** (0.0261)	0.283*** (0.0264)	0.282*** (0.0264)	0.281*** (0.0264)
Par_Edu	-0.141*** (0.0277)	-0.141*** (0.0288)	-0.138*** (0.0291)	-0.137*** (0.0292)	-0.137*** (0.0292)	Par_Edu	-0.130*** (0.0329)	-0.126*** (0.0344)	-0.119*** (0.0349)	-0.118*** (0.0349)	-0.118*** (0.0349)
Par_Occ	-0.0931*** (0.0243)	-0.0963*** (0.0252)	-0.118*** (0.0255)	-0.119*** (0.0256)	-0.113*** (0.0256)	Par_Occ	-0.160*** (0.0282)	-0.165*** (0.0294)	-0.194*** (0.0298)	-0.195*** (0.0298)	-0.189*** (0.0298)
HS_grade	0.471*** (0.0215)	0.486*** (0.0224)	0.491*** (0.0227)	0.493*** (0.0226)	0.494*** (0.0227)	HS_grade	-0.317*** (0.0255)	-0.325*** (0.0267)	-0.327*** (0.0270)	-0.324*** (0.0270)	-0.323*** (0.0270)
HS_type	0.0415* (0.0222)	0.0454** (0.0231)	0.0474** (0.0233)	0.0481** (0.0233)	0.0566** (0.0234)	HS_type	-0.340*** (0.0253)	-0.349*** (0.0263)	-0.343*** (0.0266)	-0.342*** (0.0267)	-0.333*** (0.0267)
lg_REEF		-0.0607 (0.0453)	0.0611 (0.0470)	0.0510 (0.0472)	0.0564 (0.0472)	lg_REEF		-0.0365 (0.0520)	0.0655 (0.0538)	0.0551 (0.0538)	0.0587 (0.0540)
lg_GDP			0.560*** (0.0401)	0.4645*** (0.0506)	0.2183*** (0.0522)	lg_GDP			0.428*** (0.0465)	0.322*** (0.0492)	-0.6401*** (0.0611)
lg_IND					0.2943*** (0.0371)	lg_IND					0.3067*** (0.0432)
lg_IQI				0.0215*** (0.0032)	0.0203*** (0.0032)	lg_IQI				0.0246*** (0.0040)	0.0234*** (0.0041)
Constant	1.100*** (0.0223)	1.080*** (0.0231)	-4.437*** (0.396)	-3.474*** (0.421)	0.333 (0.761)	Constant	0.298*** (0.0258)	0.268*** (0.0269)	-3.941*** (0.459)	0.261 (0.729)	1.111 (0.892)
N	59787	55087	53998	53998	53998	N	59787	55087	53998	53998	53998
Pseudo R^2	0.0374	0.0379	0.0396	0.0401	0.0409	Pseudo R^2	0.0374	0.0379	0.0396	0.0401	0.0409
Log likelihood	-58433.22	-53818.99	-52658.54	-52630.42	-52584.47	Log likelihood	-58433.22	-53818.99	-52658.54	-52630.42	-52584.47
LR chi2	4538.49	4237.21	4338.04	4415.41	4486.17	LR chi2	4538.49	4237.21	4338.04	4415.41	4486.17
Prob>chi2	0.0000	0.0000	0.0000	0.0000	0.0000	Prob>chi2	0.0000	0.0000	0.0000	0.0000	0.0000

Note: Standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 10. Multilevel Model. Estimated coefficients (0*: Defending*; 1: *Productive*)

	(1)	(2)	(3)	(4)
Gender	-0.469***	-0.462***	-0.461***	-0.460***
	(0.0130)	(0.0134)	(0.0134)	(0.0134)
Par_Edu	-0.0888***	-0.0861***	-0.0857***	-0.0843***
	(0.0171)	(0.0177)	(0.0177)	(0.0177)
Par_Occ	-0.0676***	-0.0670***	-0.0694***	-0.0706***
	(0.0150)	(0.0154)	(0.0155)	(0.0155)
HS_grade	0.295***	0.306***	0.300***	0.301***
	(0.0131)	(0.0136)	(0.0136)	(0.0136)
HS_type	0.0357***	0.0379***	0.0380***	0.0388***
	(0.0137)	(0.0141)	(0.0142)	(0.0142)
lg_REEF		0.114***	0.0300	0.0328
		(0.0321)	(0.0301)	(0.0344)
lg_GDP			0.222***	0.2173***
			(0.0250)	(0.0329)
lg_IQI				0.0091***
				(0.0225)
Constant	0.691***	0.676***	-1.508***	-1.458***
	(0.0139)	(0.0141)	(0.248)	(0.251)
N	44222	41410	41087	41087
Log likelihood	-25933.822	-24361.691	-24201.759	-24200.386
LR test	416.30	333.12	109.84	68.99
Prob	0.0000	0.0000	0.0000	0.0000

Note: Standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. If lg_IND is included among regressors, estimation does not converge.

Table 11. Multilevel Model. Estimated coefficients. (0*: Defending*; 1: *Productive*)

	(1)	(2)	(3)	(4)
Gender	-0.462***	-0.462***	-0.462***	-0.461***
	(0.0134)	(0.0134)	(0.0134)	(0.0134)
Par_Edu	-0.0866***	-0.0846***	-0.0850***	-0.0871***
	(0.0177)	(0.0177)	(0.0177)	(0.0177)
Par_Occ	-0.0672***	-0.0666***	-0.0667***	-0.0664***
	(0.0155)	(0.0155)	(0.0155)	(0.0155)
HS_grade	0.301***	0.302***	0.302***	0.301***
	(0.0136)	(0.0136)	(0.0136)	(0.0136)
HS_type	0.0395***	0.0396***	0.0395***	0.0392***
	(0.0142)	(0.0142)	(0.0142)	(0.0142)
lg_REEF	0.181***	0.102***	0.103***	0.139***
	(0.0350)	(0.0321)	(0.0330)	(0.0426)
lg_GDP	0.104***	0.149***	0.138***	0.0614**
	(0.0311)	(0.0304)	(0.0295)	(0.0253)
lg_IND	0.0994***	0.154***	0.146***	0.113***
	(0.0245)	(0.0227)	(0.0229)	(0.0288)
lg_CORR	0.00377**			
	(0.00183)			
lg_GOV		-0.00898**		
		(0.00410)		
lg_REG			0.00369	
			(0.00351)	
lg_RULE				0.0895***
				(0.0217)
Constant	0.100	-0.0701	0.0156	0.616**
	(0.371)	(0.365)	(0.352)	(0.301)
N	41087	41087	41087	41073
Log likelihood	-24182.954	-24186.639	-24188.518	-24175.382
LR test	75.99	67.91	62.29	34.04
Prob	0.0000	0.0000	0.0000	0.0000

Note: Standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. If lg_VOICE is included among regressors, estimation does not converge.

Figure 1. Returns to agents and the shares of rent seekers and defending.

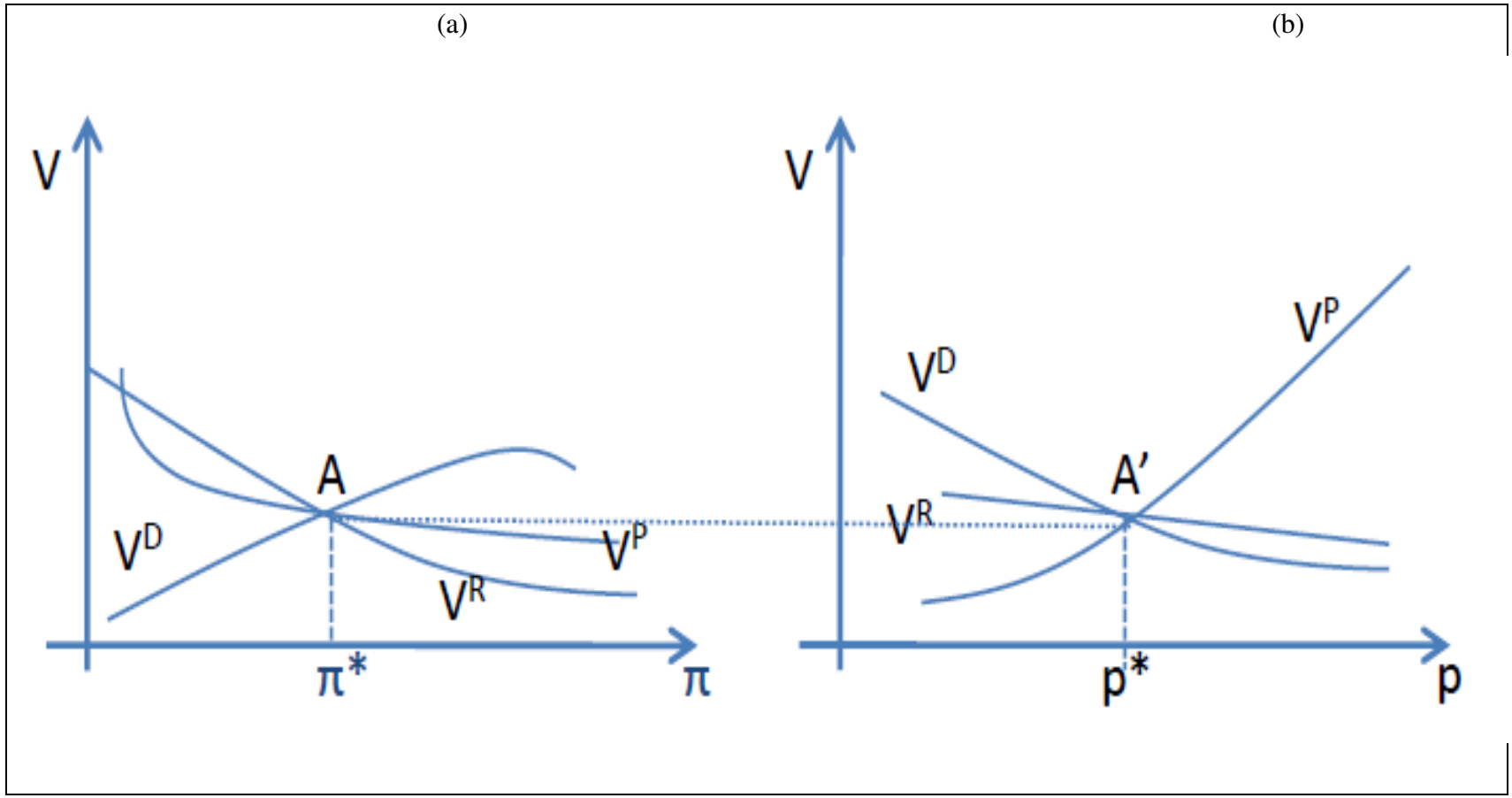

Figure 2. The effects of improving institutional quality (lower q)