QUANTITY VS. QUALITY: HOW TO READ THE INVENTOR CVs

Myriam Mariani
Bocconi University
IEP and *Cespri*, Milano, Italy
myriam.mariani@uni-bocconi.it

Marzia Romanelli
University of Pisa, Italy
romanelli@sssup.it

VERY PRELIMINARY DRAFT, Sept 2005

Comments most welcome

Abstract

This paper studies the determinants of the *quantity* versus the *quality* of the inventors' patents. The empirical investigation uses a sample of 792 inventors drawn from a new survey dataset (the PatVal-EU dataset), and the information on their EPO patents applied for during 1988-1998. The paper explores three aspect of inventors' productivity: 1) the number of EPO patents that they produce; 2) their average quality; 3) the quality of their most valuable patent. By jointly estimating the three equations we find that inventors with the highest level of education and employed in large firms invent a larger number of patents. No factor other than the number of patents invented affect instead the average quality and the maximum. This is suggestive of an increasing returns process whereby more ideas is the key input to produce better ideas.

Acknowledgments: This paper is based on the data collected through the PatVal-EU project (Contract HPV2-CT-2001-00013) funded by the European Commission. We thank Bart Verspagen for providing us with patent citations, Grid Thoma for helping us with the EPO search, and the ZEW Research Institute for giving us access to the *SearchEngine* matching software.

1. Introduction

Innovation and human capital are key factors for the growth of firms, and for economic growth more generally. Yet, little is known about the key actors of this process, the industrial inventors, and the determinants of their productivity.

Traditional contributions focus on scientists rather than industrial inventors. By using scientific publications as the measure of their output, previous studies have shown that the scientists' productivity distribution is skewed (Lotka, 1926; de Solla Price, 1963; Allison and Steward, 1974), and that vintage and age matter with scientists becoming less productive as they become older (Oster and Hamermesh, 1998, Levin and Stephan, 1991). This holds after controlling for individual fixed effects that proxy for differences in motivation and ability.

Our knowledge about industrial inventors is sparser. The difficulty to obtain information about the individual inventors prevented previous research from performing systematic empirical studies on this matter. The existing evidence is based on small scale samples, specific industries and firms (e.g. Narin and Breitzman, 1995; Ernst et al., 2000).

This paper uses information on a large sample of European inventors to study the determinants of the *quantity* and *quality* of their innovations. Inventors' productivity may assume different aspects. While the number of patents that they produce is one of them, the inventors often acquire visibility for the "value" of their innovations, and sometimes their reputation depends on one or few major innovations (Jones, 2005). This calls for an indicator of the quality (value) of their patents. We therefore follow Lanjouw and Schankerman (2004) and extract a common component from several patent indicators (forward citations, backward citations, number of claims, family size) to construct a composite index that proxies for the technological and economic importance of the innovations. We then study the determinants of the inventors' productivity measured in three ways:

1) Number of patents that the inventors invented, and that were applied for at the EPO in 1988-1998;
2) Average importance of these innovations as measured by the common component indicator;
3) Expected maximum value of the patents invented by the individual inventor, i.e. the inventor's patents with the highest level of the indicator.

We use a sample composed of 792 European inventors. Information on their individual characteristics is drawn from the PatVal-EU survey, which interviewed the inventors of 9,017 EPO patents applied for in 1993-1997. The empirical investigation uses all the EPO patents applied for in 1988-1998 in which our 792 individuals appeared as inventors. We jointly estimate three equations at the inventor-level with 1)-3) above as dependent variables. Our controls are individual, firm, industry, and country characteristics.

Our empirical results suggest an intriguing story about the drivers of inventors' productivity. Once we control for countries and sectors, the number of inventors' patents increases with age, academic degree, employment in a large firm, and the number of patents of the organization in which the inventors are employed. Surprisingly, none of these factors produce a direct effect on the value (both average and maximum) of the innovations.

Further investigation, however, reveals that the average and the maximum value increase as the number of the inventor's patents increases. In other words, inventors with more ideas also have better ideas. While this is natural in the case of the maximum as the expected value of an order statistics increases with the number of trials (e.g. Mood *et al*, 1974), it is more interesting in the case of the average quality, as it suggests no regression to the mean in the invention process at the level of the individual inventor. In both cases, however, it is notable that there are only indirect effects of inventor and firm characteristics arising through the number of the inventor's patents. This implies a sort of hierarchy in the effects that we are studying. Individual characteristics or other factors affect the number of patents that the inventors produce, and this, in turn, affects (positively) their quality.

To identify the effect of the number of patents on the expected average and maximum quality of the innovations we did not make restrictions on the observed factors affecting the three measures of productivity. We retrieve the effect of the number of patents on the two quality variables from assumptions about the structure of the variance-covariance matrix of the residuals of the three equations.

The rest of the paper is organised as follows. Section 2 briefly discusses the background literature. Section 3 presents the data. Sections 4 and 5 describe the estimation procedure and show the results of the empirical tests. The final section summarizes and concludes the paper.

2. Background literature

The determinants of research productivity over the researchers' life cycle have been studied in the economic literature as well as in other disciplines. A pioneer work is Lotka (1926) who shows that research productivity is highly concentrated among scientists. Others confirm these findings and explain them with differences in the distribution of ability among scientists, and with the allocation of recognition and resources to the most productive individuals that makes them even more productive – the so-called Matthew Effect (Merton 1968; Allison and Steward, 1974; Cole 1979; David, 1994).

Other authors show that vintage and age matter in many disciplines with older scientists becoming less productive (Dalton and Thompson, 1971; Goldberg and Shenhav, 1984). Oster and Hamermesh (1998), for example, follow the careers of 208 economists in the economic departments of 17 top research institutions who received Ph.D degrees between 1959 and 1983. They provide evidence that publishing diminishes with age. They also show that there is persistent heterogeneity. The most productive economists early in their careers keep on producing high-quality research (though at a slower rate) as they become older. Levin and Stephan (1991) examine the research productivity of scientists over their life cycle in six scientific areas. They find that scientists become less productive over time.[1]

The existing evidence about industrial inventors is based on small-scale samples, specific industries and firms. Narin and Breitzman (1995) tested the Lotka's inverse square law of productivity on a sample of inventors in the R&D departments of four companies in the semiconductor industry. Similarly, Ernst *et al.* (2000) studied the research productivity of the inventors of 43 German companies, both in terms of quantity and quality of their innovations (see also Ernst, 1998 for a study at the firm level). In general, this literature confirms the skewness of the productivity

[1] Cole (1979) however finds that age is curvilinearly related to the quantity and quality of scientists' productivity in cross-sectional analysis, and that there is no relationship between age and productivity when using longitudinal data on the scientists' career.

distribution among researchers but, given the difficulties to obtain individual specific information and to trace the careers of the industrial inventors, it does not test the reasons behind these disparities.

Our work contributes to deepen the study of the relationship between inventors' research productivity, age and other determinants of the *quantity* versus *quality* of the innovations. We measure quality by a common component indicator developed by Lanjouw and Schankerman (2004). In fact there are quite a few contributions that demonstrate that there is a positive relationship between patent indicators and the actual ex-post value of the innovations as given by traditional accounting evaluation (Hall et al., 2005).[2] For example, Trajtenberg (1990) shows that there is a close association between patent counts weighted by forward citations and the social value of innovations in the Computer Tomography Scanner industry. Harhoff *et al.* (1999) demonstrate that the number of backward citations to other patents and to non-patent literature, and the number of citations received after the publication of a patent are positively correlated with the value of the innovation. This holds also for patents applied in many countries, and for patents that incur in opposition and annulment procedures (Harhoff and Reitzig, 2004). Schankerman and Pakes (1986) use patent renewal data to estimate the value of patent rights. Others use patent family size (i.e. the number of countries in which the patent is asked for protection) and the number of claims in the patent application as indicators of the value of the patents (see, for example, Putnam, 1996).

Lanjouw and Schankerman (2004) use multiple patent indicators to construct a composite index that substantially reduces the measured variance in patent quality. Gambardella *et al.* (2005) also show that this composite indicator is highly correlated with the monetary value of the patents.

3. Our dataset

3.1. Data source

Our source of data is the PatVal-EU database. PatVal-EU was constructed from a survey conducted in 2003-2004, which interviewed the inventors of 9,017 patents granted by the EPO with priority date in 1993-1997, and located in France, Germany, Italy, the Netherlands, Spain and the United Kingdom (for details and descriptive statistics see Giuri et al, 2005; and European Commission, 2005). The PatVal-EU database provides critical information for the purpose of our study about the age, education, career, affiliation, and the motivations to invent of a fairly large sample of EPO inventors.

To analyse the productivity of these inventors we collected additional information on the number and quality of the innovations that they produced over their career. Since this task was extremely time-consuming, we started with a sub-sample of 792 PatVal-EU inventors. This sub-sample was selected by taking all the German, Italian, Dutch and English inventors that responded to the PatVal-EU questionnaire on patents invented in five technological classes: Information Technology, Optics, Biotechnology, Chemical Engineering, and Civil Engineering.

---

[2] On the limitations of patent indicators see Griliches, 1990; Almeida and Kogut, 1999; Singh, 2005; Alcacer and Gittelman, 2004.

We allocated patents to technological classes by employing the ISI-INIPI-OST classification.[3] We then selected one micro ISI-INIPI-OST class from each of the five macro-classes, and we checked the distribution of the patents belonging to the five micro-classes compared to the whole PatVal-EU sample. Information Technology, Chemical Engineering, Civil Engineering, Optics, and Biotechnology well represent the average characteristics of the inventors and the innovation process of the macro-classses in which they are classified (respectively, Electrical Engineering, Process Engineering, Mechanical Engineering, Instruments, Chemicals and Pharmaceuticals). Specifically, the Wald tests on the mean difference of some key variables between countries and technologies confirm that our sample is not biased in any particular direction. Only in Biotechnology the share of inventors with a PhD is higher than the average. In Optics the estimated economic value of the innovations is higher than the average of the macro-class.

For the 792 inventors in our sample we complemented the PatVal-EU data by data on each inventor's career. We employed the database Delphion to collect all the patents that the inventors invented and that were either applied or granted by the EPO in 1988-1998.[4] Although we could only take this window because EPO data are not fully reliable before end of the 80s, an eleven-year window is large enough to capture a sizable portion of the inventors' career. The search on Delphion was followed by a matching procedure to assign each patent to the specific inventor and solve problems of homonymy. Indeed, quite a few inventors had the same first and last names of some inventor in our list, but they were not the same inventors.[5] To eliminate homonymous inventors we employed a matching software that run on two tables: a searching table with information on our 792 PatVal-EU inventors, and a reference table with all the potential EPO matching patents extracted from Delphion. The match between the EPO patents and the inventors was performed by using a set of weighted criteria: last inventor's name, first name, second name, technological class. The matching software browsed the reference table, and reported the "probability for each patent to be a good match" with the inventor in the searching table. At the end of this process, we further checked the list of inventors manually for the presence of homonymous inventors. The PatVal-EU data (like the name of the applicant, the technological class, the extent of inventors' mobility) and some investigation on the web helped solve the doubtful cases.

In the end we obtained data on 4,396 patents invented by our 792 PatVal-EU inventors. For each patent we have information on the innovation like the number of claims, the number of states in which the innovations was patented, the name and location of the applicant organisation, the IPC classes in which the patent was classified, and the number of forward and backward citations. We also obtained data on the inventor that contributed to develop the patent, and the organization in

[3] The ISI-INIPI-OST classification was developed by the German Fraunhofer Institute of Systems and Innovation Research (ISI) together with the French patent office (INIPI) and the Observatoire des Science and des Techniques (OST). It is based on the International Patent Classification (IPC) and distinguishes among 5 macro and 30 micro classes.

[4] Delphion is an on-line database released and managed by Thomson Corporation that collects all the European patent applications published by the EPO and issued since 1979. We downloaded all the European patents, either applied or granted, in which the last and first name of the inventors corresponded to one of our 792 inventors.

[5] The matching software is *SearchEngine v. 5.751*, which is an experimental software developed at the ZEW research institute by Thorsten Doherr. In our sample we still have the problem of "misspelled names" (i.e. inventors whose name is misspelled in the patent document). We plan to solve this problem by using the matching software together with a Soundex algorithm.

which he was employed, as collected from the PatVal-EU survey. Table 1 describes the composition of our sample of 792 inventors who contributed to invent the 4,396 patents.

[TABLE 1]

The share of women in the sample is extremely low: it is larger in Biotechnology, and smaller in Civil and Chemical Engineering, with no significant differences across countries. The average age of the inventors is 45, with some variation across countries and technologies. Moreover, the more a technology is science-intensive, the larger is the share of inventors with Ph.Ds as in Optics and Biotechnology compared to Civil Engineering, and to the overall share of 33.4%. In this respect, Italy deserves attention: only 5.1% of the inventors in the sample have a Ph.D degree. Table 1 also shows that the average number of patents per inventor in our database is 5.5, with a peak for German and Italian inventors (7.4 and 6.4 respectively).

3.2. Productivity measures and regressors

As we mentioned in the Introduction, the aim of the paper is to study the determinants of the inventors' research productivity in terms of the *quantity* and *quality* of the innovations that they produce.

We use the number of patents of the inventors as a measure of the quantity of their research output (NX_i). Our quality measures are obtained from Lanjouw and Schankerman's (2004) composite quality-adjusted patent index, which proxies for both the technological and economic impact of the innovations. As they discuss, the index reduces the variance in patent quality compared to employing only one of the traditional value indicators (see Hall *et al*, 2001, for a survey). Moreover, since changes in the separate indicators that compose the index might stem from factors other than the actual changes in the technological or economic impact of the innovations, the advantage of the index is that it is cleaned from differences among patents that depend on country, time and technological characteristics. Finally, Gambardella *et al.* (2005) show that the monetary value of patents is highly correlated with a common component indicator similar to the one that we employ in this paper.

The common factor is the unobserved characteristic of a patented innovation that influences all the four following indicators:

Forward Citations: a patent must cite all related prior patents, and the patent examiner eventually checks and changes them to ensure that all appropriate citations are included. These citations identify the right of the patentee, and they are a signal for the technological importance of a patent as a source of knowledge on which subsequent patents are built. We obtained the number of patents that cite the 4,396 patents in our sample, and we isolated the number of citations received within 5 years of the patent publication date.

Backward Citations: we also collected the number of prior patents cited by the 4,396 patents. Backward citations are an indicator of others working on similar research fields, and therefore they signal the importance of the technological area.

Claims: a claim describes the features of the invention, and it defines the property rights protected by the patent. The inventor and the patent applicant have an incentive to write as many claims as they can, but the examiner may require some of them to be dropped. The larger is the number of claims, the broader and the greater is the expected profitability of an innovation.

Family Size: this is the number of countries in which the innovation is asked for patent protection. Since the extension of the patent rights in each country is costly, the larger the number of countries, the higher the expected value of the innovation.

From our 4,396 patent sample we retrieved the parameters estimates to construct the index. To do so we controlled for some observed characteristics of the patents: the nationality of the inventors, the application year, and the primary micro-technological class in which the patents are classified. For each indicator k (k = 1,...4) of the pth patent we run the following multiple-indicator model with one latent common factor:

$$y_{kp} = \beta_0 + \boldsymbol{\beta}'\boldsymbol{x}_p + \lambda_k q_p + \varepsilon_{kp}$$

where y_{kp} indicates the value of the kth indicator for the pth patent (in logs). The common factor is q with factor loadings λ_k, while x_p denotes the vector of observed controls. To solve the overidentification problem (see Lanjow and Schankerman, 2004, for details) we use the structure of the correlation between the measurement errors of the 4 equations that suggests that the parameter estimate of Family size is close to zero and it is not significant. We therefore set the correlation between the measurement error in the Family Size equation and the measurement errors in the equations of the other indicators equal to zero. The top part of Table 2 shows the correlation between the errors of the four equations. The bottom part shows the parameter estimates of the indicators that compose the quality index.

[TABLE 2]

The quality index developed in this way was then employed to obtain two inventor-level measures of the patent quality. One of them is QX_i, which is the average value of the index across all the patents of each inventor.

However, inventors often acquire visibility for one or few major inventions (Jones, 2005; Zucker *et al.*, 1998). Moreover, valuable patents are rare and lie in the very left-end side of the patent value distribution. We therefore decided to look at the factors that explain the probability of inventing “the most” valuable innovation amongst those produced by the individual inventors. Accordingly, our second quality measure is MX_i, which is the “best” innovation amongst the inventor’s patents. This is measured by the inventor’s highest value of the common component indicator across his patents.

Figures 1 and 2 show the distribution of NXi, QXi, and MXi across the 792 inventors. Consistently with other contributions on these topics, Figures 1 and 2 confirm that the distribution of the three productivity measures is skewed with few inventors being very productive.

[FIGURE 1 and 2]

The three productivity measures are the dependent variables of our regressions. The regressors are inventors characteristics, characteristics of the organisation in which they work, country and technological dummies. Table 3 lists and defines the variables.

[TABLE 3]

Age, sex, education and motivation to invent are the inventor level characteristics that we included in the regressions. We use five age classes for the age of the inventors (AGE1-AGE5).

We employ four types of inventors' education degree: Secondary School (SecSc), High School (HighSc), University BA or Master (Uni); Ph.D (PhD). We expect the level of education being positively correlated with the research performance of the inventors. Not only is this because education proxies for the inventors' "exogenous" ability, but also because education might be a signal that inventors use to search for a good "match" between their research potential and the characteristics of the employer organisation that better exploits such potential.

As far as their motivations to invent are concerned, the PatVal-EU survey asked the inventors to use a scale from 1 to 5 to rate the importance of different rewards from patenting. In particular, we considered the following three rewards: economic compensation (COMP), reputation (REP) and career advances (CAREER). We expect that "love for research" and reputation compared to "love for money" and career advances have a different impact on the expected quantity versus quality of the innovations. (See, for example, Levin and Stephan (1991) who make the argument of "investment motivated" vs. "consumption motivated" research).

Firm size and the number of patents the organisation was granted in the PatVal-EU database are the two firm characteristics we include in the regressions.[6] We read the number of patents the organisation was granted (PATSORG) as a proxy for its propensity to patent. As far as the type of organisation and its size is concerned, large firms (LARGE) have the financial resources to produce and apply for patent protection for a larger number of innovations than smaller companies (SME). Large firms are also endowed with a large pool of heterogeneous and specialised researchers that are involved in a large number of research projects. We expect these characteristics to positively affect the probability of inventing important innovations. However, it is difficult to say anything with respect to the average quality of such innovations in large companies. It is possible that the financial strength to apply for patent protection not only for important innovations, but also for less valuable ones might lead to a lower average quality.

FAM is the average number of interconnected patents to which the inventor's innovations belong. A patent can either stand alone, or it can be connected to other patents that are listed in the patent document. We use FAM as a proxy for the extent to which the inventors produce patents that are part of groups of related patents, which, in turn, might be the cause of a large NXi.

[6] We used information on the type of employer organisation as it was given by our PatVal-EU inventors. To check whether the inventor changed employer during 1988-1998 we checked whether the application years of the inventor's patents in our sample were included in the period between the year in which the inventor was hired by the PatVal-EU organisation and the year in which he left this employer. Since almost all our inventors did not change job while developing the patents in our sample, we use information on the employer as given by the PatVal-EU database. The next version of this paper will collect information on the different employers of the very few cases of "mobile" inventors.

Finally, we control in all the regressions for the country of the inventors (COUNTRY) and for the macro technological classes in which the patents are classified (TECH).[7] Table 4 provides the descriptive statistics of the variables.

[TABLE 4]

Figures 3, 4, 5 and 6 sketch the relationship between age, gender, academic degree, type of organisation in which the inventor is employed on the one hand, and their average productivity on the other hand. It might be worth keeping in mind that the quality index is depurated from sectoral, country and time effects.

[FIGURES 3-4-5-6]

Figures 3 and 5 suggest that age and education are positively correlate with NXi, while it seems they have no relationship with QXi and MXi. Surprisingly, males seem to be better than women in producing innovations (Figure 4). Most likely being a man is a proxy for the time and effort spent on doing research. Finally, private companies, and in particular large firms have an advantage in doing a large number of innovations, even though they are not particularly better than medium and small firms and universities in doing more valuable innovations on average (Figure 6). Only MXi is higher in large firms. The next two Sections will combine all this information together by means of multivariate regression analysis, and will derive the net impact of each variable on the inventors' productivity.

4. Specification and Estimation: Step 1

We start by estimating a set of three equations, one for each productivity measure (NX_i, QX_i and MX_i) by means of Seemingly Unrelated Regressions (SUR).

$$\begin{cases} NX_i = \boldsymbol{x}_i'\boldsymbol{\alpha}_1 + \varepsilon_{1i} \\ QX_i = \boldsymbol{x}_i'\boldsymbol{\alpha}_2 + \varepsilon_{2i} \\ MX_i = \boldsymbol{x}_i'\boldsymbol{\alpha}_3 + \varepsilon_{3i} \end{cases}$$

The subscript i denotes the inventor, while α_1 α_2 α_3 are the coefficients to be estimated for the impact of the x_i organisation and inventors' characteristics on the three productivity variables. All the variables are in logarithms. Table 5 presents the estimated results. Since NX_i is a count variable, we also show the results of this equation by using a Negative Binomial regression.

[TABLE 5]

[7] In the case of inventors with patents in different macro-classes we use the sector where the majority of his patents falls.

The general picture we draw from the estimates in Table 5 is that inventors' and firm specific factors produce a different impact on the number of patents per inventor compared to the average and maximum expected quality of the innovations. Specifically, they are positively correlated with the probability of producing a large number of innovations (NXi), but they loose power when they enter the quality equations (QXi and MXi).

This is the case of AGE: as the inventors become older, the probability that they produce a large number of patents rises.[8] The same AGE variable, however, does not affect the probability of inventing valuable innovations (both average and maximum). Also the academic degree of the inventors produces a positive effect on NXi: inventors with a university or master BA, and in particular inventors with a Ph.D have a higher probability of producing a large number of patents. Again, however, the inventors' academic degree is not correlated with the quality indicators. Being a MALE is also positively and significantly correlated with NXi and MXi. This is due, probably, to the very low number of women in our sample (which is a signal anyway!), and to the fact that, on average, male inventors can put more effort and time than women in their job. Once we control for all these factors, none of the three motivations to invent affects the inventors' productivity.

As expected, being in a large firm that applies for a large number of patents matters for NXi. The dummy LARGE is positive and statistically significant at 0.05 on the number of patents; PATSORG is positive and statistically significant at 0.01.[9] These results are confirmed by the Negative Binomial estimation in the right-end column of Table 5. Firm characteristics (i.e. LARGE and PATSORGS) are positive and statistically significant also in the MXi equation. By contrast, none of the factors we include in the regressions matter for QXi.

This is puzzling, and in particular it is surprising that: first, inventors' personal characteristics are not correlated with any of the quality indexes; and second, that firm characteristics matter only in the NXi and MXi equations. If this is true, the story that arises is one in which some structural firm and personal characteristics influence only the quantity of the innovations that the inventors produce, while serendipity and stochastic factors explain the probability of inventing high quality innovations. In particular inventors' personal characteristics do not matter for the expected value of the innovations. To understand this story better we chose to go one step further. In particular we check our hypothesis that the scale of the research performed by the inventors produces an impact on the expected quality of their innovations. The next section tests this hypothesis.

---

[8] We are aware of the potential source of selectivity biases in our paper. As discussed by Cole (1979) and Levin and Stephan (1991), age and ability are likely to be positively correlated, as productive inventors are more likely to remain in the sample as they age. The next version of the paper will deal with this problem by using the Olsen (1980) correction to predict the likelihood that the inventors are active in the innovation business.

[9] As robustness checks we include the variables "age" and "age^2" in place of the age classes. Age is positively and significantly correlated with the probability of inventing NXi, while age^2 has a negative impact (that is statistically significant in the Negative Binomial estimation). We also include the variable MOBILITY for inventors who changed employer at least once after the patent in the PatVal-EU database, but the estimated coefficient is never significantly correlated with the inventors' productivity. Finally, when PATSORG is removed from the regressions, the coefficient of LARGE gains significance. We also estimated the three equations in our system by using the number of citations received in the 5 years after the publication date as the dependent variables (the total number, the average, and the maximum by inventor) in place of the quality index. Not surprisingly given the importance of forward citation in the common component indicator, the results do not differ significantly from those in Table 5.

5. Specification and Estimation: Step 2

This section builds empirical evidence for a story that we envision about the drivers of inventors' productivity. The story goes as follows. More skilled inventors are employed in large companies that have the financial resources and the research capabilities to develop a large number of innovations. This is suggested by the estimated results of the NXi equation. The productivity based on the number of innovations listed in the inventors' CV is also valuable for the organisations in which the inventors are employed. Hoisl (2005) reports two examples of inventors who are rewarded according to the number of patents they contributed to invent. This is the case of Siemens AG that in 2004 honoured 13 inventors responsible for about 600 patents invented in the same year. Similarly, the WIPO Award scheme launched in 1979 grants prizes to the inventors for their "outstanding research activities and numerous patented inventions".

We will go one step further by showing that not only is the number of innovations per inventor important for the inventors' CV and recognition, but it also affects the expected quality of his innovations. In other words, quantity leads to better quality: skilled inventors are employed in large firms that give the inventors the opportunity to produce a large number of innovations and to apply for patents. In turn, since "ability [is] also a product of learning" (Griliches, 1970), this wide exploration process in research leads to a higher expected quality of the innovations.

To find empirical evidence consistent with this story we make an assumption about how NX_i enters in the QX_i and MX_i equations. Specifically:

$$\begin{cases} NX_i = \boldsymbol{x}_i'\boldsymbol{\alpha}_1 + \varepsilon_{1i} \\ QX_i = \boldsymbol{x}_i'\boldsymbol{\alpha}_2 + \vartheta_2 NX_i + \varepsilon_{2i} \\ MX_i = \boldsymbol{x}_i'\boldsymbol{\alpha}_3 + \vartheta_3 NX_i + \varepsilon_{3i} \end{cases}$$

By substituting the NX_i equation in the QX_i and MX_i equations, the system above can be rewritten as follows:

$$\begin{cases} NX_i = \boldsymbol{x}_i'\boldsymbol{\alpha}_1 + \varepsilon_{1i} \\ QX_i = (\boldsymbol{\alpha}_2 + \vartheta_2\boldsymbol{\alpha}_1)\boldsymbol{x}_i' + \vartheta_2\varepsilon_{1i} + \varepsilon_{2i} \\ MX_i = (\boldsymbol{\alpha}_3 + \vartheta_3\boldsymbol{\alpha}_1)\boldsymbol{x}_i' + \vartheta_3\varepsilon_{1i} + \varepsilon_{3i} \end{cases}$$

To retrieve θ_2 and θ_3 we need something that affects NX_i without entering the QX_i and MX_i equations but, unfortunately, we do not have a structural model that justifies such inclusion and exclusion restrictions. A possible solution is to estimate θ_2 and θ_3 from the variance-covariance matrix between the residuals of our three equations. To do so it is natural to let the errors ε_2 and ε_3 of the two quality equations to be correlated between each other but not with ε_1.[10] We then propose a structure in which NX_i enters the QX_i and MX_i equations, but not vice-versa. In the end we obtained the estimates of θ_2 and θ_3 and their standard errors as shown in Table 6.

[10] This assumes that we have in fact been able to include all the observed factors that affect NX and MX. Although this is clearly an assumption, it is also true that we control for many individual and institutional factors in our regressions.

[TABLE 6]

The sign and statistical significance of θ_2 and θ_3 confirm our story that the larger the number of innovations the inventor produces, the higher the probability to invent more valuable patents on average, and to produce a technological hit. It is also interesting that, at the inventor level, there is no regression to the mean in the quality of the innovations: θ_2 is positive and statistically significant. Moreover, the Wald test for the α_2 and α_3 parameters of LARGE, PATORG, DEG3 and DEG4 confirms the statistical insignificant effect of these variables on both QX_i and MX_i once NX_i is included in the two quality equations. The insignificance of these variables on MX_i is particularly noteworthy. It suggests that the earlier significance of these variables in the reduced form system estimated in Section 3 (Step 1) was proxying for the impact of NX_i. Put differently, these variables do not seem to affect directly the distribution from which the inventors draw their inventions. They affect NX_i, which in turn affects the expected value of the maximum.

All in all, our story is that inventors with the highest level of education and employed in large firms invent a large number of patents. This starts an increasing returns process where more ideas means also better ideas: as the number of patents the inventor produces increases, we find that the expected average and the maximum value of his innovations increase as well. Up to this point, however, we cannot say whether this is the result of a statistical process where the larger the number of draws, the higher the probability to get a success, or whether this is the output of some learning process through which knowledge accumulates and raises the probability to invent something valuable.

To provide additional evidence on the role played by the inventors' personal characteristics on the probability to be hired by a large firm, we estimated a Probit regression where the dependent variable is 1 if the employer organisation is a large firm (LARGE) and 0 if it is a small or medium enterprise (SME). We exclude from the sample the inventors employed in Universities, Government organisations and other institutions. Inventors' characteristics (age, academic degree, gender, motivations to invent, family size), country and technology dummies are the regressors. Table 7 shows the estimated results.

[TABLE 7]

As expected, the level of education matters for the inventors to be hired by a large company. The positive coefficients of Uni and PhD are statistically significant at 0.01 and 0.05 respectively, confirming that large firms employ more skilled inventors, and that inventors might use the level of education as a signal for their "ability" to the potential employer. This is consistent with a recent contribution by Baumol (2005) who argues that large-firm research and development laboratories require highly educated personnel. He also provides the example of Procter & Gamble (one of the world's largest holders of US and global patents) that has developed a global research organisation with over 7,500 scientists, including 1,250 PhDs.[11]

[11] Moore (1911) argues that "Large establishments are able to carry out the work of selection [of more capable individuals] because in consequence of their large capital and better organisation, they offer opportunities for more capable individuals to reap the reward of their differential ability". Since then there is the idea that larger firms create jobs that match with more productive individuals. For example, Idson and Oi (1999) confirm that "the adoption of

6. Conclusions

This paper studied the drivers of inventors' productivity. It estimated the impact of individual characteristics and organisation factors on the expected *quantity* versus *quality* of the innovations produced by the European inventors.

We used information on the characteristics of 792 inventors. Our data are drawn from the PatVal-EU survey conducted in 2003-2004 by interviewing the inventors of 9,017 European patents. For the 792 sample inventors we also downloaded 4,396 patents that they contributed to invented and that were applied at the EPO in 1988-1998.

Quantity is simply measured by the number of EPO patents the inventors applied in the period 1988-1998. We considered two measures of patent quality: the average importance of the patents produced by each inventor in 1988-1998, and the importance of the most valuable patent amongst each inventor's patents. In both cases, quality is proxied by a common component index as developed by Lanjouw and Schankerman (2004).

We regressed these three types of inventors' productivity on individual characteristics (i.e. age, academic degree, motivation to invent, gender) and on the characteristics of the organisation in which the 792 inventors in our sample are employed (large firm, small and medium enterprises, universities and other public research organisations), including the propensity of the organisation to apply for patent protection. We employed seemingly unrelated regressions to estimate the effect of each variable on the probability to develop a large quantity of innovations as compared to produce high quality innovations. We also used a method based on the matrix of variance-covariance between the residuals of the three equations to retrieve the estimated impact of the number of innovations developed by the inventor on the average and maximum quality of such innovations.

Our results indicate that the drivers of quantity are different from those that impact on quality. Specifically, the higher the level of education of the inventor, the higher the probability of being employed in a large firm. Better-educated inventors in large (innovative) firms are more likely to produce a large number of innovations. This is probably because the number of innovations is a signal to the firm in order to identify and reward productive inventors. It is also because the large firm provides the inventors with the resources needed to develop and apply for many patents. The inventors' characteristics, however, do not impact directly on the expected value of the innovations (both average and maximum).

Their impact, however, is indirect through the number of innovations that skilled inventors produce. We find that the number of patents developed by the inventors is positively and significantly correlated with the average and maximum quality of their patents. In other words, more ideas lead to better ideas. Interestingly, our results also suggest that there is no regression to the mean at the inventor level: as the number of patents per inventor gets larger, the average quality of the innovations raises as well.

Whether this is the output of a probabilistic process where the larger the number of trials, the higher the expected quality, or whether this is because the inventors that produce many patents also explore a large number of technological opportunities, enter in many different research projects,

advanced technologies, employment of inherently more able individuals, and higher work standards go together to raise labor productivity [...]".

interact with many researchers, and undertake a learning process that increases the probability of producing better innovations, we don't know, yet. Whatever it is, however, not only is the distribution of inventors' productivity skewed across individuals, but also quantity is correlated with quality, and with the inventors' personal characteristics. This results reinforce the "retaining the best" policy that Narin and Breitzman (1995) put forward: if ability is concentrated in a few key individuals, the organisations in which they are employed (being there private companies or public research laboratories) should make their best to retain them. This is most important if these individuals are responsible for both the quantity and quality of the organisation's innovations, and if, at the inventor level, there is no regression to the mean in the average value of the innovations.

7. References

Alcacer, J., Gittleman, M. (2004), How do I Know what you Know? Patent Examiners and the Generation of Patent Citations, *New York University*, Working Paper

Allison P.D., Stewart J.A. (1974), Productivity Differences Among Scientists: Evidence for Accumulative Advantage, *American Sociological Review*, Vol. 39, No. 4, pp. 596-606

Almeida, P., Kogut, B. (1999), Localization of Knowledge and the Mobility of Engineers in Regional Networks, *Management Science*, Vol. 45, No. 7, pp.905-917

Baumol, W.J. (2005), Education and Innovation: Entrepreneurial Breakthroughs Versus Corporate Incremental Improvements, A.B. Jaffe, J. Lerner and S. Stern (Ed.), *Innovation Policy and the Economy*, NBER, 33-56

Cole, S. (1979), Age and Scientific Performance, *American Journal of Sociology*, Vol. 84, 958-77

Dalton G.W., Thompson, P.H. (1971), Accelerating Obsolescence of Older Engineers, *Harvard Business Review* 49: 57-67

David P., (1994), Positive Feedbacks and Research Productivity in Science: Reopening Another Black Box, in Grandstrand O. (eds), *Economics of Technology*, North Holland, 65-85

De Solla Price, D. (1963), *Little Science, Big Science*, Yale University Press, New Haven, CT

Ernst H., (1998) Industrial Research as a Source of Important Patents, *Research Policy*, 27(1): 1-15

Ernst H., Leptien C., Vitt, J. (2000), Inventors are Not Alike: The Distribution of Patenting Output Among Industrial R&D Personnel, *IEEE Transactions on Engineering Management* 47(2): 184-199

European Commission (2005), *The Value of European Patents. Evidence from a Survey of European Inventors*, Final Report of the PatVal-EU Project

Gambardella A., Harhoff D., Verspagen B., (2005) The value of Patents, paper presented at the NBER Conference, The Economics of Intellectual Properties, Cambridge (MA) July, 19th 2005

Giuri P., Mariani M., Brusoni S., Crespi G., Francoz D., Gambardella A., Garcia-Fontes W., Geuna A., Gonzales R., Harhoff D., Hoisl K., Lebas C., Luzzi A., Magazzini L., Nesta L., Nomaler O., Palomeras N., Patel P., Romanelli M., Verspagen B. (2005), *Everything you always wanted to know about inventors (but never asked): Evidence from the PatVal-EU survey*, LEM Working Paper No. 2005/20, Sant'Anna School of Advanced Studies, Pisa.

Griliches Z. (1970), Notes on the Role of Education in Production Functions and Growth Accounting, L. Hansen (Ed.). *Education, Income, and Human Capital*, Cambridge, NBER, 71-115

Griliches Z. (1990), Patent Statistics as Economic Indicators: A Survey, *Journal of Economic Literature*, Vol. 28, 1661-1707

Goldberg A.I., Shenhav, Y.A. (1984), R&D Career Paths: Their Relation to Work Goals and Productivity, *IEEE Transactions on Engineering Management* 31(3): 111-117

Hall B., Jaffe A., Trajtenberg M. (2001), The NBER Patent Citations Data File: Lessons, Insights and Methodological Tools, *NBER working paper* No. 8498

Hall, B.H., Jaffe, A., Trajtenberg, M. (2005), Market value and patent citations, *RAND Journal of Economics*, forthcoming

Harhoff D., Narin F., Scherer F.M., Vopel K. (1999), Citation Frequency and the Value of Patented Innovation, *Review of Economics and Statistics*, Vol. 81, 511-515

Harhoff, D., Reitzig, M. (2004). Determinants of opposition against EPO patent grants — the case of biotechnology and pharmaceuticals. *International Journal of Industrial Organization* 22, 443–480

Hoisl, K., (2005), A Closer Look at Inventor Productivity, LMU Discussion Paper

Jones B. F. (2005), Age and Great Invention, *NBER Working Paper* 11359

Lanjouw J.O., Schankerman M. (2004), Patent Quality and Research Productivity: Measuring Innovation with Multiple Indicators, *Economic Journal*, No.114, pp. 441-465

Levin S.G., Stephan, P.E. (1991), Research Productivity over the Life Cycle: Evidence for Academic Scientist, *American Economic Review*, Vol. 81, No. 1, pp.114-32

Lotka, A.J. (1926), The Frequency Distribution of Scientific Productivity, *Journal of the Washington Academy of Science* 16(2): 317-323

Merton, R.K. (1968), The Matthew Effect in Science, *Science* 159 (January): 56-63

Mood A., Graybill F.A., Boes D.C., (1974)Introduction to the Theory of Statistics, McGraw-Hill

Narin, F., Breitzman, A. (1995), Inventive Productivity, *Research Policy* 24(4): 507-519

Olsen R., (1980), A Least Squares Correction for Selectivity Bias, *Econometrica*, 48 (7) 1815-1820

Oster S. M., Hamermesh D. S., (1998), Aging and Productivity among Economists, *The Review of Economics and Statistics*, 154-156

Putnam, J. (1996). *The value of international patent rights*, Ph.D. thesis, Yale University

Schankerman, M. and Pakes, A. (1986). Estimates of the value of patent rights in European countries during the post-1950 period, *Economic Journal*, vol. 96, pp. 1052–76

Singh, J., (2005), Collaborative Networks as Determinants of Knowledge Diffusion Patterns, *Management Science*, forthcoming

Trajtenberg, M. (1990), A Penny for Your Quotes: Patent Citations and the Value of Innovations, *RAND Journal of Economics*, Vol. 21, No. 1, pp. 172-187

Zucker, L., Darby, M., Armstrong, J., 1998. Geographically localized knowledge: spillovers or markets? Economic Inquiry 36, 65–86

Figures and Tables

Table 1. Sex, age, education and patents applied by the 792 inventors (1988-1998). Distribution by technological class and country.

	% of female inventors	average age of the inventors*	% of inventors with university BA or Master	% of inventors with Ph.D	average # of patents in 1988-1998
Information Technology (145)	1.4	41 (9.5)	51.2	33.3	5.9 (9.3)
Optics (139)	5.0	42 (8.9)	34.1	46.4	6.4 (6.9)
Biotechnology (53)	9.8	43 (9.0)	15.4	48.1	4.6 (4.2)
Chemical Engineering (197)	1.0	46 (9.8)	39.8	38.3	6.1 (8.5)
Civil Engineering (258)	0.8	48 (9.1)	44.3	19.6	4.6 (5.8)
Germany (305)	2.9	47 (9.6)	46.9	33.4	7.4 (8.5)
Italy (119)	2.3	44 (10.8)	56.4	5.1	6.4 (10.3)
The Netherlands (160)	1.7	42 (7.5)	13.8	54.4	3.5 (3.9)
UK (208)	1.9	45 (10.1)	43.4	34.6	3.8 (4.3)
Total	2.3	45 (9.7)	40.6	33.4	5.5 (7.4)

* The age of the inventors in calculated as 1995-date of birth. Standard deviations in parenthesis.
Note: The number of observations is in parenthesis next to the technological classes and countries' names.

Table 2. Correlation between the errors of the four equations and parameter estimates.

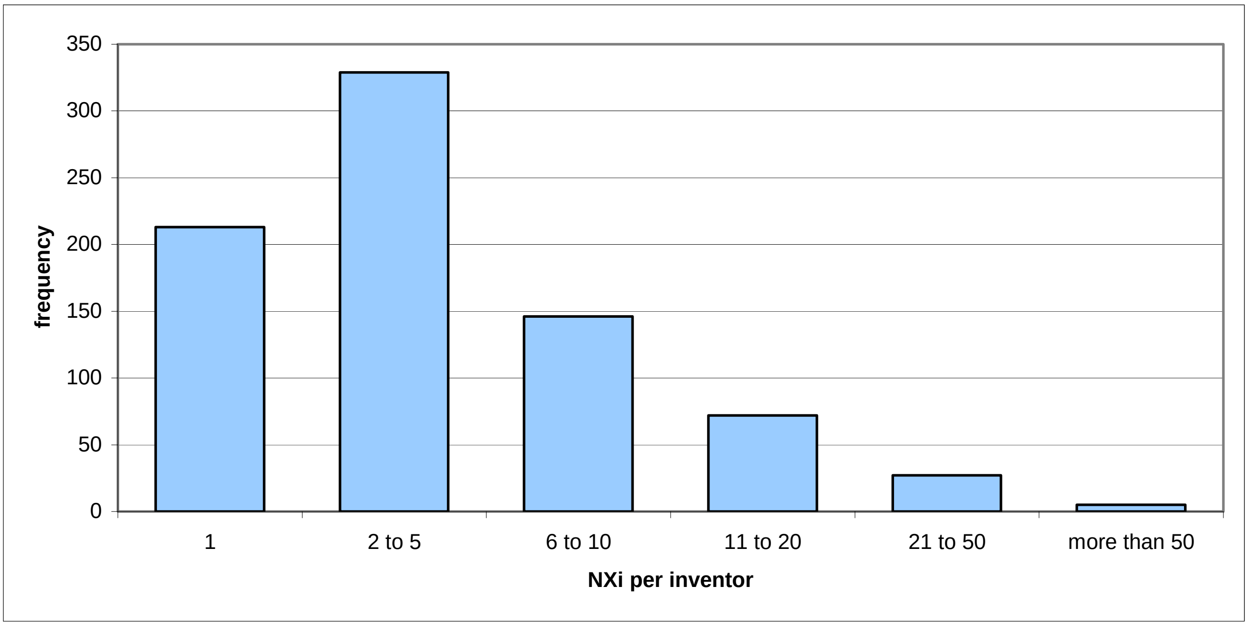
Errors of the 4 equations	Forward citations equation	Backward citations equation	Family size equation	Claims equation
Forward citations eq.	1.000			
Backward citations eq.	0.321***	1.000		
Family size eq.	0.013	0.018	1.000	
Claims eq.	0.050***	0.027*	-0.008	1.000

Parameter Estimates of the Common Factor model

Variable (log)	
Forward citations (5 years after publication date)	0.34*** (0.09)
Backward citations	0.22*** (0.06)
Number of claims	0.06*** (0.02)

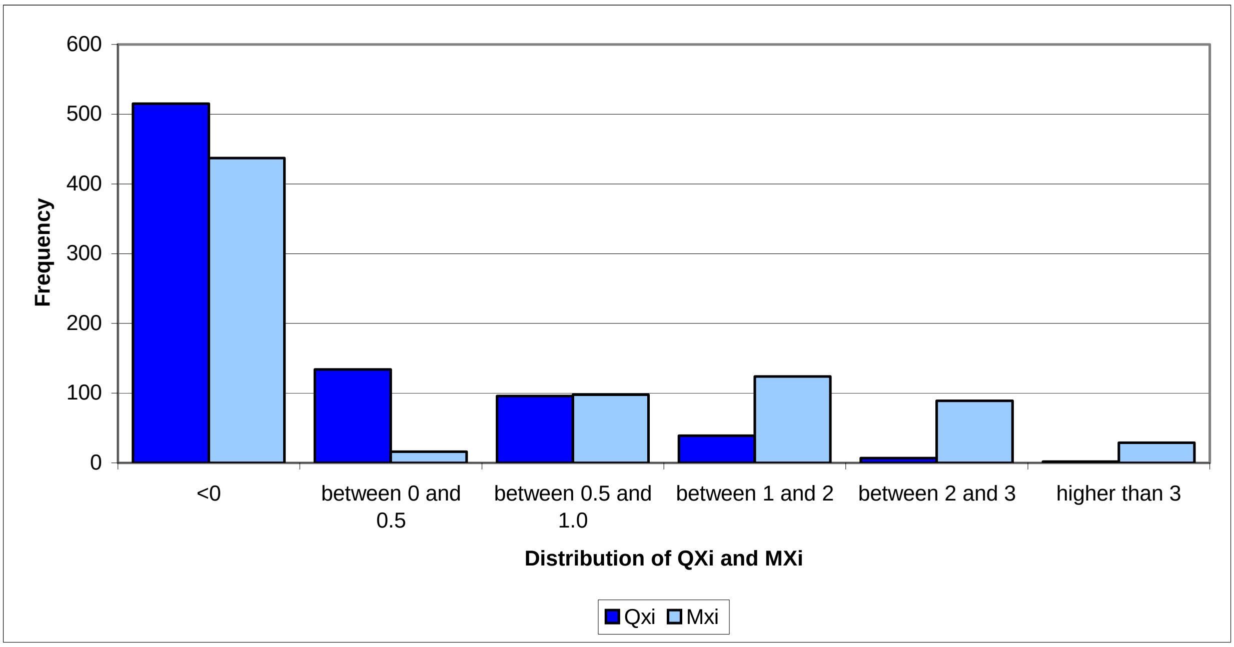
Note: Nationality, Year, Technological dummies are included. Standard errors in parenthesis.
Significant at: *0.1 level; **0.05 level; ***0.01 level.

Figure 1: Distribution of NXi.



Note: # obs. 792

Figure 1: Distribution of QXi and MXi.



Note: # obs.: 792. QXi and MXi in logs.

Table 3. List of variables.

	Productivity measures
NXi	Number of patents invented by the inventor and applied or granted in 1988-1998 (application date) (Source: EPO)
QXi	Average quality of the patents invented by the inventor (application date 1988-1998) as measured by the mean of the common component index across patents. (Source: EPO)
MXi	Maximum quality of the patents invented by the inventor (application date 1988-1998) as measured by the most valuable (i.e. highest common component index) of the individual inventor's patents. (Source: EPO)
	Inventors characteristics
AGE	Dummy. Age of the inventors defined as 1995-date of birth. Each inventor is classified as: under 30 (AGE1); 31-40 (AGE2); 41-50 (AGE3); 51-60 (AGE4); over 60 (AGE5). AGE1 is the reference group in the regressions. (Source: PatVal-EU)
DEGREE	Dummy. Highest Academic degree of the inventor at the time in which he developed the PatVal-EU innovation. We grouped them as follows: secondary school or lower (SecSc); high school (HighSc); University BA or Master (Uni); Ph.D (PhD). The reference group is SecSc (Source: PatVal-EU)
GENDER	Dummy. Male or Female inventor (Source: PatVal-EU)
COMP, REP, CAREER	The inventors were asked to rate the importance of different rewards from the patent by using a scale from 1 to 5. We created a dichotomous variable with value 1 if the inventor assigned 4 or 5 to the importance of the specific reward; 0 if the importance is below 4. We use the following three rewards: Economic compensation (COMP); Reputation (REP); Career Advances (CAREER) (Source: PatVal-EU)
FAM	Average number of patents that compose the group of interrelated patents (family) of which the innovation is part (Source: EPO). Average size of the group by inventor.
	Characteristics of the employer organisation
EMPL	Dummies. Type of employer organisation: Large firm (more than 250 employees), Medium firm (between 100 and 250 employees), Small firm (less than 100 employees), University; Goverment Institutions; Other. We use the following dummies: Large firm (LARGE), Medium and Small firm together (SME); Universities and Government institutions (GOV). GOV is the reference group. (Source: PatVal-EU)
PATSORG	Number of patents granted to the employer organisation (Source: PatVal-EU)
	Controls
TECH	Dummies for the macro ISI-INIPI-OST technological classes in which the inventor's patents are classified: Electrical Engineering (ELENG) (reference group), Process Engineering (PRENG), Mechanical Engineering (MECENG), Instruments (INST), Chemicals & Pharmaceuticals (CHEM).
COUNTRY	Country of the inventor: Germany (DE), Italy (IT), the Netherlands (NL), UK (UK) (reference group)

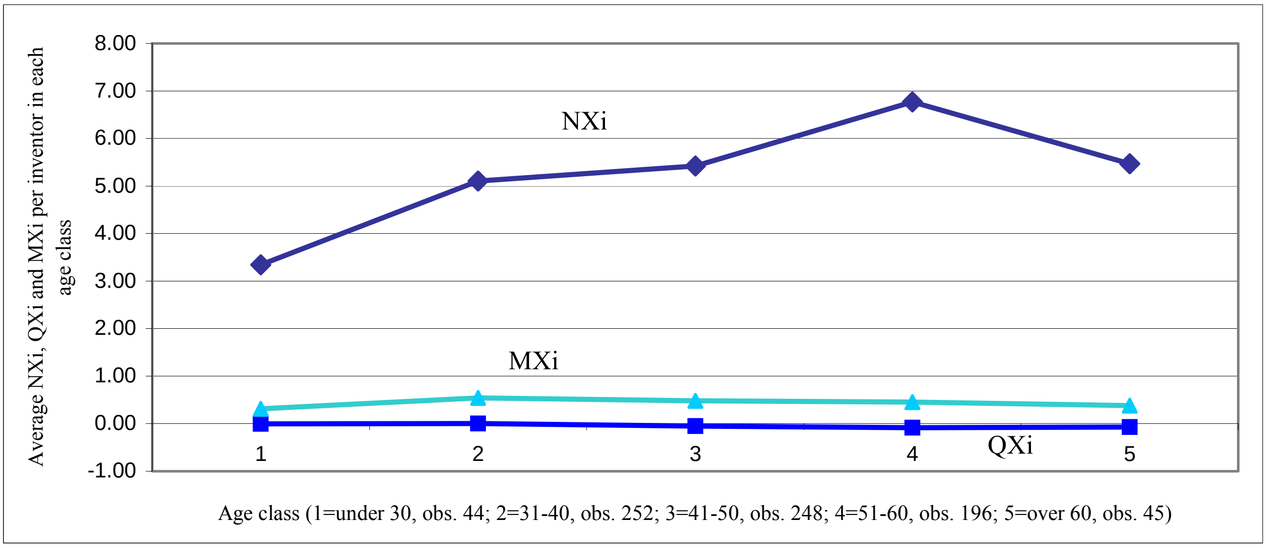
Source: PatVal-EU dataset and EPO

Table 4. Descriptive statistics.

	Mean	S.D.	Min.	Max
	Productivity measures			
NXi	5.53	7.40	1	87
QXi (in logs)	-0.04	0.64	-0.57	5.25
MXi (in logs)	0.49	1.18	-0.56	5.25
	Inventors characteristics			
AGE	2.93	1.01	1	5
DEGREE	3.26	0.82	1	4
GENDER	0.98	0.15	0	1
COMP	0.39	0.49	0	1
REP	0.36	0.48	0	1
CAREER	0.54	0.50	0	1
FAM	10.27	4,42	3	22
	Characteristics of the employer organisation			
LARGE	0.60	0.49	0	1
SME	0.24	0.43	0	1
GOV	0.11	0.32	0	1
PATSORG	33.44	72.95	1	286

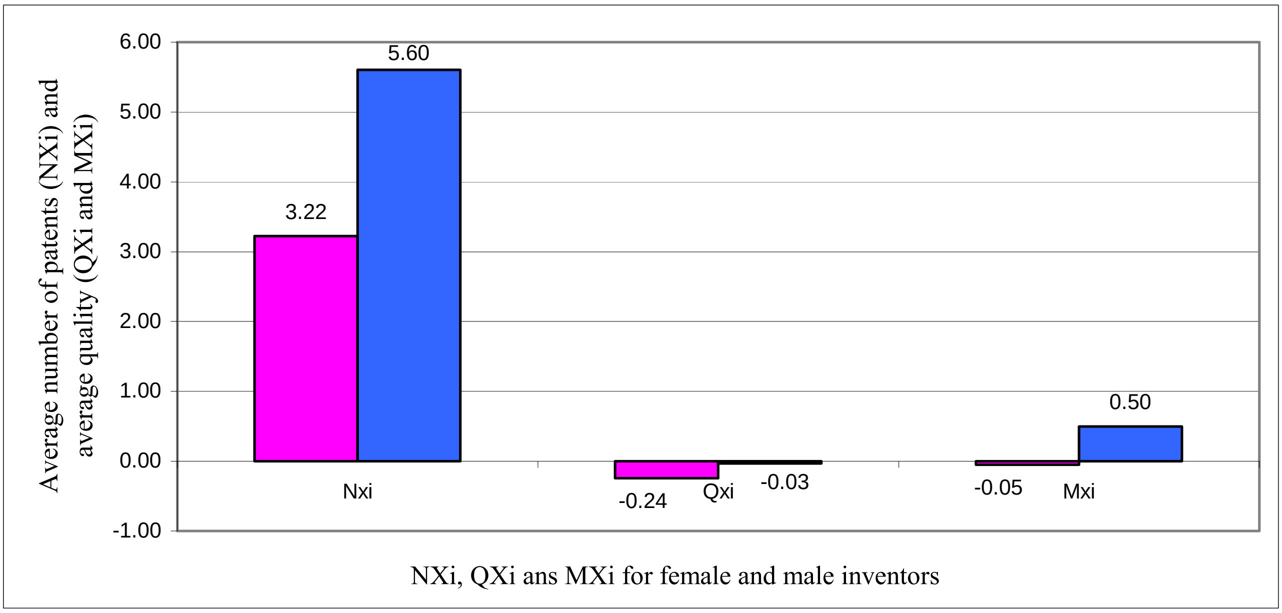
Source: PatVal-EU dataset and EPO

Figure 3: Productivity and Age. Average NXi, QXi and MXi by age class.



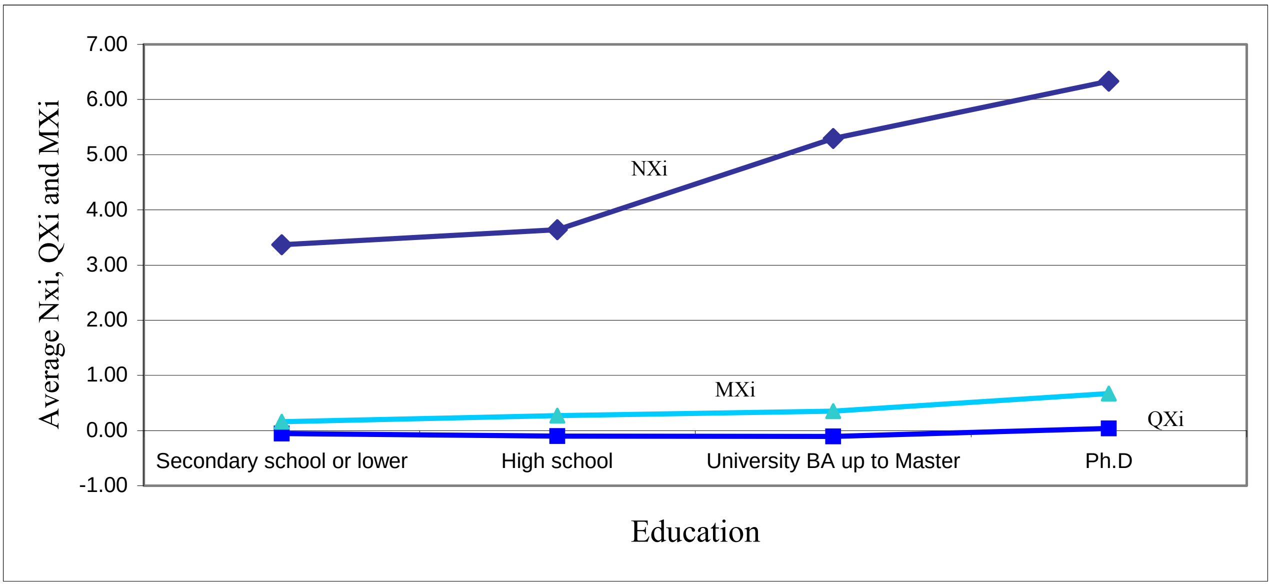
Note: # obs.= 785. QXi and MXi in logs.

Figure 4: Productivity and Gender. Average NXi, QXi and MXi by male and female inventors.



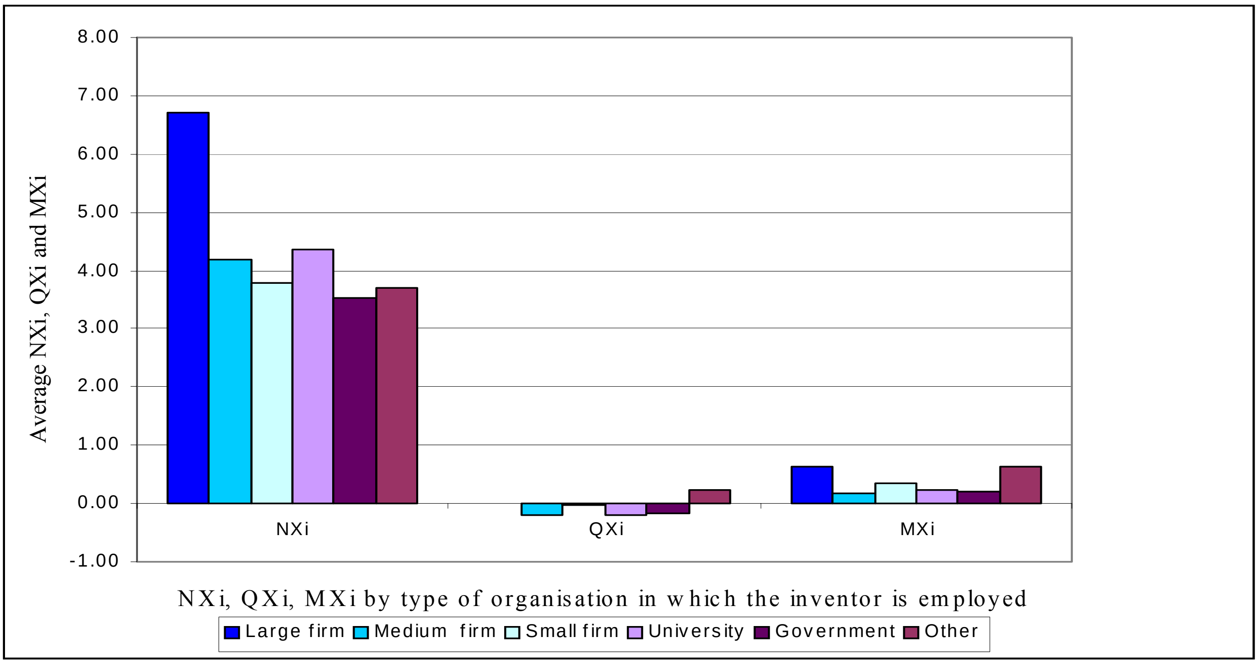
Note: # obs: 785 (18 female and 767 male inventors). QXi and MXi in logs.

Figure 5: Productivity and Academic Degree. Average NXi, QXi and MXi by academic degree classes.



Note: # obs.: 785 (Secondary School: 38; High School: 72; University BA or MA: 319; Ph.D: 356). QXi and MXi in logs.

Figure 6: Productivity and type of employer organisation Average NXi, QXi and MXi by type of organisation.



Note: # obs.: 721 (Large firm: 459; Medium Firm: 67; Small Firm: 115; University: 46; Government: 24; Other: 10). QXi and MXi in logs.

Table 5. Estimates of SUR and Negative Binomial Regression. Variables in logs.

Dependent variables: Productivity measures				
	SUR			NegBin
	NXi	QXi	MXi	NXi
AGE2	0.26* (0.14)	-0.03 (0.13)	0.16 (0.18)	0.36** (0.16)
AGE3	0.37** (0.14)	-0.03 (0.13)	0.24 (0.19)	0.50*** (0.17)
AGE4	0.49*** (0.15)	-0.001 (0.13)	0.23 (0.20)	0.57*** (0.18)
AGE5	0.42** (0.21)	-0.04 (0.15)	0.25 (0.28)	0.57** (0.22)
DEGREE: HighSc	0.04 (0.19)	-0.11 (0.17)	0.006 (0.23)	-0.09 (0.23)
DEGREE: Uni	0.29* (0.17)	-0.07 (0.15)	0.17 (0.19)	0.32 (0.20)
DEGREE: PhD	0.55*** (0.18)	-0.06 (0.15)	0.34 (0.20)	0.68*** (0.20)
MALE	0.55** (0.25)	0.20 (0.15)	0.65*** (0.24)	0.52* (0.27)
COMP	-0.01 (0.08)	0.01 (0.05)	-0.03 (0.10)	-0.04 (0.08)
REP	0.09 (0.08)	-0.02 (0.06)	0.02 (0.11)	0.14 (0.09)
CAREER	0.05 (0.08)	-0.001 (0.05)	-0.01 (0.10)	-0.02 (0.08)
FAM	0.04 (0.09)	0.07 (0.06)	0.12 (0.11)	0.11 (0.09)
LARGE	0.30** (0.12)	0.15* (0.09)	0.36** (0.16)	0.28** (0.13)
SME	0.04 (0.12)	0.05 (0.09)	0.12 (0.16)	0.02 (0.13)
PATSORG	0.08*** (0.02)	0.02 (0.015)	0.06** (0.03)	0.09*** (0.02)
Const.	-0.86 (0.39)	-0.58 (0.28)	-1.39 (0.46)	-0.63 (0.43)
# obs. 646				
Log Likelihood	–1884.88			-1730.96

Note: Robust Standard Errors in parenthesis. Sample: 792 inventors. All regressions include dummies for inventors' country and macro ISI-INIPI-OST technological class.
Coefficient significant at: *0.1 level; **0.05 level; ***0.01 level.
For the Negative Binomial regression: $\alpha = 0.55$ (0.04); R-squared = 0.20

Table 6. Parameter estimates: θ_2 and θ_3.

Variable (log)	
θ_2	0.07** (0.03)
θ_3	0.57*** (0.05)

Note: Robust Standard Errors in parenthesis.
Coefficient significant at: *0.1 level; **0.05 level; ***0.01 level.

Table 7. Probit Estimates of the probability to be hired by a large firm.

Dependent variables: Employment in a Large firm = 1	
AGE2	0.00 (0.84)
AGE3	0.04 (0.31)
AGE4	0.26 (0.32)
AGE5	-0.56 (0.38)
HighSc	0.22 (0.34)
Uni	0.73*** (0.28)
PhD	0.66** (0.29)
MALE	-0.001 (0.63)
COMP	-0.08 (0.15)
REP	0.23 (0.16)
CAREER	0.19 (0.14)
FAM	-0.001 (0.16)
Const.	0.02 (0.84)

Note: Variables in logs. Standard Errors in parenthesis.
Sample: 542 inventors employed in Large firm (406) and SME (136).
All regressions include dummies for inventors' country and macro ISI-INIPI-OST technological class.
Coefficient significant at: *0.1 level; **0.05 level; ***0.01 level.
Log likelihood = -258.214; R-squared = 0.17