A Closer Look at Inventor Productivity – What Makes the Difference?

Karin Hoisl*
University of Munich

Draft Version – August 2005

Abstract
Previous work on the productivity of scientists and engineers reveals that productivity is highly concentrated within a small number of researchers. To design an efficient reward scheme, firms have to identify their key inventors. To help identifying these inventors, this paper analyses the determinants of inventor productivity. To do so, at first, an improved productivity measure is created which controls for changes in patent propensity over time and across different technical areas. Additionally, the appropriateness of whole patent counts versus fractional patent counts is discussed. In a second step, the productivity measure is used to identify determinants of inventor productivity. Whereas former research looked on determinants related either to the inventor or the inventive environment, this study is the first to integrate both types of determinants. Results show that the level of education has a strong influence on inventor productivity. Making use of external sources of knowledge also increases productivity. In particular exploiting the knowledge from other patent documents and from competitors increases inventive output. Finally, firm size has an important impact on productivity.

Acknowledgement
The survey data used in the analysis originates from a coordinated survey carried out in Italy, France, Spain, the Netherlands, the United Kingdom and Germany. Financial support from the European Commission within the PatVal-Project is gratefully acknowledged.

* *Karin Hoisl*, Ludwig-Maximilians-University Munich, Insitute for Innovation Research, Technology Management and Entrepreneurship, Kaulbachstr. 45, 80539 Munich, Tel: ++49 (0)89 2180 5626, e-mail: hoisl@bwl.uni-muenchen.de.

1 Introduction

In December 2004, the German producer SIEMENS AG honored its top inventors for the tenth year in a row. The 13 winners were responsible for about 600 inventions made in 2004, accounting for approximately 7.5% of all reported inventions in the same year. The SIEMENS Inventor's Award is aimed at enhancing inventive and innovative activities of the 50,000 employees in the field of R&D. In 2004, the electronics group spent 5.1 billion Euros in R&D (accounting for 6.8% of sales), resulting in 23.2 million Euros per working day.[1] Another example for an award to motivate inventors is the WIPO Award scheme launched in 1979. This scheme aims at improving the image of inventors through recognition of their merits and also at promoting inventive activities. Between 1979 and 2000, 561 prizes were awarded to inventors from 77 countries. The WIPO does not interfere in the selection of the nominees. Independent national or international organizations nominate potential candidates to be selected by an inventor's award committee, composed of representatives of government authorities, academia, associations of inventors, chambers of commerce, and industry. "Outstanding research activities and numerous patented inventions" are given as criteria for the decision to grant a prize.[2]

Previous work on the productivity of scientists and engineers reveals that productivity is highly concentrated within a small number of researchers. Treating all inventors equally leads to a decrease of the motivation of the most productive inventors. To design an efficient incentive system, such as the two examples described above, firms have to identify their key inventors. To help identifying these inventors, this paper analyses the determinants of inventor productivity. To do so, at first, an improved productivity measure is created. The literature, so far, has not distinguished between discrete or complex product industries or industries which changed their patenting behavior. To measure productivity, patents have been added up, regardless of the priority year or technical area. In this context, Hall (2004) proposes a strategic shift of the patenting behavior of U.S. firms in certain industries, resulting in a "patent explosion". This paper improves on the current literature by developing an adjusted measure for productivity. A productivity index is applied, controlling for patent propensity changes over time and across different technical areas. Additionally, the appropriateness of whole patent counts versus fractional patent counts[3] to measure productivity is discussed.

In a second step, the improved productivity measure is used to identify determinants of inventor productivity. Whereas former research looked on determinants related either to the inventor or the inventive environment, this study is the first to integrate both types of determinants. To detect the relevant determinants, this paper uses survey data on 3,049 German inventors, who hold at least one granted European patent. The inventors were requested to provide demographic information as well as information on the R&D process underlying their patented invention. To trace the patent counts of each inventor over time, the EPOLINE® database of the European Patent Office was used to search for all patents

[1] See http://www.siemens.com/index.jsp?sdc_p=cz3s5uo1233409pnfl0mi1033645&sdc_sid=33336152884&sdc_bcpath=1026937.s_5%2C&.

[2] See http://www.wipo.int/innovation/en/wipo_awards/invention.htm.

[3] Fractional patent count means that a fraction of a patent is assigned to an inventor with respect to the number of co-inventors.

belonging to the 3,049 inventors with priority dates between 1977[4] and 1999, resulting in a total of 29,971 EP patents.

Results show that the level of education has a strong influence on inventor productivity. Making use of external sources of knowledge also increases productivity. In particular exploiting the knowledge from other patent documents and from competitors increases inventive output. Finally, firm size has an important impact on productivity.

The remainder of this paper is divided into five sections. The following section provides an overview of the theoretical and empirical literature on inventor productivity. The third section proposes hypotheses derived from existing literature. Section 4 contains the description of the data used in this paper. In section 5 descriptive statistics will be provided. Furthermore, a log linear regression will be estimated to identify determinants of productivity differences. Finally, section 6 discusses the results and provides implications for further research.

2 Theoretical Background and Empirical Evidence

Previous work on inventor productivity focused on the analysis of productivity distributions using pure patent or quality adjusted patent counts as a proxy for productivity. In 1926, Lotka examined the number of name entries appearing in the decennial index of Chemical Abstracts between 1907 and 1916 as well as the names appearing in the index of Auerbach's "Geschichtstafeln der Physik". Plotting the logarithmic number of persons responsible for one, two, or more contributions against the logarithmic number of contributions resulted in an almost linear relationship with a gradient of approximately two. Based on these results, Lotka formulated the "inverse square law of productivity". According to Lotka's Law, the number of researchers producing n scientific contributions is proportional to $1/n^2$. Assigned to the example of Chemical Abstracts, this means that when 100 scientists appear exactly once in the index of Chemical Abstracts, it would be a share of $1/2^2$ (= 25) who would appear twice within the same time period, and so on. Price (1965) showed that productivity is highly concentrated within a small number of key scientists. He proposed the "square root law of elitism" suggesting that a scientific community in a particular research field contains an elite group of scientists, almost identical to the square root of all community members. This elite group is responsible for about 50 percent of the entire scientific output within this research field.

Narin and Breitzman (1995) were the first to apply Lotka's Law to the distribution of patents. They studied the output of inventors in R&D departments of four semiconductor firms. Their findings suggest that a relatively small number of key inventors is responsible for ten or more patents whereas a large number of inventors are responsible for only one patent. The top 1% of inventors is five to ten times as productive (in number of patents) as the average inventor. The top decile of the inventors is at least three to four times as productive as the average inventor (Narin/Breitzmann 1995). A second study, confirming the results of Narin and Breitzman, was conducted by Ernst et al. (2000). The authors use a sample of 43 German companies, active in the chemical, electrical, and mechanical engineering industry. Ernst et al. use quality-weighted patent counts as a measure for inventor productivity. Results show that

[4] Although the first European patent application was filed on June 1, 1978 with the European Patent Office, priority dates from 1977 are found in the dataset. A priority from 1977 occurs when an application claims priority under the Paris Convention from a counterpart application filed less than one year earlier in another country; such an application has a priority date equal to that of the earlier application. See http://www.patents.com/patents.htm#priority.

"key inventors are characterized by a large number of patent applications which are of high quality" (Ernst et al. 2000: 184).

Another line of research tried to analyze the determinants of productivity differences between researchers. Two hypotheses are proposed by Allison and Steward (1974): the Sacred Spark Hypothesis and the Accumulative Advantage Hypothesis. According to the Sacred Spark Hypothesis, differences in productive capacity arise due to substantial, predetermined dissimilarities among scientists. Cole and Cole (1973), for instance, state that scientists differ distinctly in their ability and motivation for doing research. The Accumulative Advantage Hypothesis, on the contrary, provides a generalization of the Matthew Effect, which can be described as follows: "the accruing of greater increments of recognition for particular scientific contributions to scientists of considerable reputation and the withholding of such recognition from scientist who have not yet made their mark" (Merton 1968). Allison and Stewart (1974) propose that the allocation of recognition and resources make highly productive scientists even more productive or lead at least to the maintenance of output productivity. Another empirical study conducted by Arora, David, and Gambardella (1998) raised the question about the effect of reputation on the volume of academic outputs. Past research publication performance is found to have an important effect on the expected amount of research funding and therefore, on future publication productivity (Arora et al. 1998).

Overall, economic studies agree on the fact that the productivity distribution among scientists is highly skew. The literature also provides initial evidence that quantitative productivity is not negatively correlated with qualitative productivity. Researchers, so far, do not agree on the reasons for the disparities in research productivity. Allison and Stewart (1974) argue that the disparity is due to differences in the assignment of resources and recognition. Liu (1986) find that match quality[5] has a positive impact on productivity. The relationship between age and productivity has been analyzed in a number of studies. Dalton and Thompson (1971), for instance, find a maximum productivity at the age of 40. To shed more light on this controversial discussion, this paper aims at analyzing productivity differences more closely using an improved productivity measure. Most notably, this paper combines the above mentioned determinants in one analysis, allowing for interactions between different determinants.

3 Hypotheses

Shockley (1957) proposes that productivity is affected by many "mental factors", such as the ability to detect important problems, technical skills and persistence. In the past, a large number of authors considered the dependence between education and ability, especially the appropriateness of education as a proxy for ability.[6] Guilford (1968) suggests regarding intellectual ability of a person "as a somewhat generalized skill that has developed through the circumstances of experience, within a certain culture, and that can be further developed by means of the right kind of experience" (Guilford 1968: 619). Griliches (1970) proposes that "ability is the product of 'learning', even if it is not all a product of 'schooling'" (Griliches 1970: 93). The author further suggests to "confess ignorance" and define ability as gross

[5] According to Topel and Ward (1992), labor mobility can be seen as a process whereby workers are sorted into jobs where their productivity is high. Widerstedt (1998) proposed that mobility is the outcome of a search and sorting process. An individual learns about his comparative advantage by trying different kinds of jobs. As a measure for match quality, differences in wages, not explained by observable factors like education or labor market experience, may be employed (Widerstedt 1998).

[6] See Becker (1964) and Denison (1964) for a survey of the relevant literature.

output of the schooling system (Griliches 1970). This paper, according to the existing literature, measures intellectual ability by using the level of education of the inventors. The following relationship is expected:

H.1: *Inventors with a higher level of education tend to show a higher productivity compared to inventors with a lower level of education.*

Inventor mobility may also provide an explanation for productivity differences.

- inter-firm mobility
  A major assumption of this so-called job match approach is that employees differ in their productivity across different jobs (Jovanovic 1979). Liu (1986) uses data derived from a labor force survey conducted among employees in the manufacturing sector of Singapore in 1974. His empirical findings propose that inter-firm and intra-firm mobility lead to a better match quality between employee and the employing company. An increase in match quality also raises the productivity of employees (Liu 1986). Topel and Ward (1992) define labor mobility as a process whereby workers are sorted into jobs where their productivity is high (Topel/Ward 1992). The matching approach supports the perception that mobility can improve inventor productivity.

- cross-country mobility
  "Brain drain", which refers to the departure of highly skilled professionals from one country or technological field for another to improve income or living conditions (Kwok/Leland 1982), may weaken the competitiveness of science and technology in the departed country, while strengthening competitors abroad. Johnson and Regets (1998) have introduced the term "brain circulation" to describe a stay abroad for a limited period of time. Due to knowledge spillovers from colleagues abroad, "brain circulation" can lead to an increase of the knowledge base at home, i.e., to "brain gain". This may increase the inventor productivity.

The following hypotheses are proposed concerning the different types of inventor mobility:

H.2: *Inventors who were mobile between firms tend to show a higher productivity compared to inventors who did not change their employer.*

H.3: *Inventors who temporarily stayed abroad tend to show a higher inventive productivity compared to inventors who stayed in their home country.*

Beyond mobility, external sources of knowledge can positively influence inventor productivity. Possible sources of knowledge are patent literature, customers, and competitors. The literature provides evidence that knowledge transfer from different sources spurs innovation. Patent documents allow inventors not only to catch up on the state-of-the-art but also to collect relevant research information. Los and Verspagen (2003) characterize patent documents as a "potential source of 'idea-creating' knowledge spillovers" (Los/Verspagen 2003: 3). Allen (1977), von Hippel (1988) and Freeman (1991) highlight the importance of customers and competitors regarding the innovativeness of firms. These external sources of information play an important role in information sharing (Afuah 2000).

The literature described above analyzes the influence of knowledge transfer on innovative output at the firm level. However, the results should also apply to the inventor level. Using different sources of knowledge should enable inventors to increase their inventive output. It is, therefore, assumed that

H.4a: *Inventors making use of patent literature, customers' knowledge or competitors' knowledge are more productive than inventors who do not use these external sources of knowledge.*

Additional external sources of knowledge are university research and scientific literature. Jaffe (1989) analyzes the effects of knowledge transfer between university research and research labs. Using data on 29 U.S. states for the years 1972-1977, 1979, and 1984, he finds a positive relationship between corporate lab patenting and university research in three technical fields: drugs, chemicals and electronics. At the inventor level, however, university research and scientific literature should not in principle spur innovation. Allen (1977) compares nineteen parallel R&D projects to analyze characteristics, distinguishing engineers from scientists. Two of them are scientific projects, the remaining 17 are technological projects. Results show that scientists receive ideas from the literature, whereas engineers hardly use scientific literature and rather employ customers or suppliers as external sources of knowledge (Allen 1977).

A possible explanation for these differences provides the concept of "absorptive capacity" (Cohen/Levinthal 1989, 1990). Absorptive capacity - the ability of a firm to recognize the value of this external information, to assimilate and to apply it to commercial R&D - is required to profit from scientific spillovers. Applying the concept of Cohen and Levinthal to the inventor level, it can be assumed that inventors generate knowledge (= inventions) by using external sources of knowledge (= e.g. scientific literature). The inventors' absorptive capacity determines the extent to which the external knowledge can be assimilated and employed. Absorptive capacity in turn depends on the extent to which the inventor is used to employ scientific sources of knowledge. It is, therefore, assumed that inventors who did doctoral or postdoctoral studies are more able to benefit from scientific research. Figure 1 illustrates the proposed dependencies.

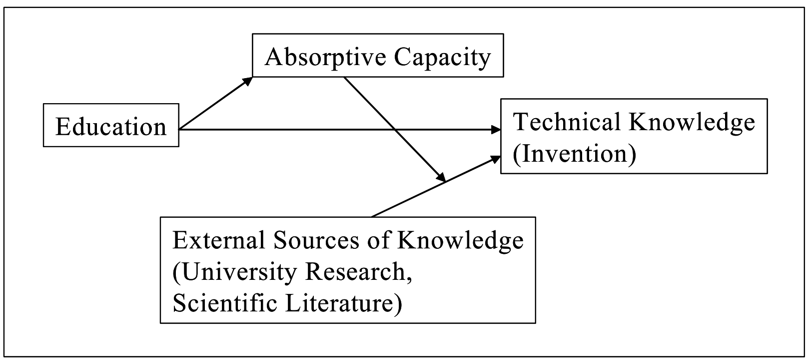

Figure 1: Absorptive capacity (figure according to Cohen/Levinthal 1990)

The following relationship is proposed:

H.4b: *Inventors who conducted scientific research (doctoral studies) increased their productivity more by using university research or scientific literature than inventors who did not conduct scientific research (i.e., the effect of university research and scientific literature on productivity should become more positive or less negative).*

Idson and Oli (1999) propose a positive relationship between labor productivity and firm size because large firms are generally early adopters of new technology. Additionally, they have more resources at their disposal to hire and retain high quality researchers. Kim et al. (2004) use longitudinal worker-firm matched data in the semiconductor and pharmaceutical industries. In both industries the authors find that inventor productivity increases with firm size. Research expenditures, sales and number of employees were used as alternative size measures. Based on the results of the existing literature, the following hypothesis is proposed:

H.5: *Inventors who are employed with a large firm tend to show a higher productivity than inventors working at small firms.*

4 Data Source and Description of the Variables

4.1 Description of the Data

The data of this survey was collected in the course of the European project PatVal, sponsored by the European Commission. The project aims at creating a database of patent characteristics obtained from a survey of European inventors named in EPO patents and from information drawn from the patent documents. Units of observation are inventors who lived in Germany at the time of application of the respective patent. 10,500 EP patents listing inventors living in Germany were chosen by a stratified random sample based on a list of all granted EP patents with priority dates between 1993 and 1997 (15,595 EP patents). A stratified random sample was used in order to oversample potentially important patents. The sample of 10,500 patents hence includes all opposed patents (1,048) and patents which were not opposed but received at least one citation (5,333), and a random sample of 4,119 patents drawn from the remaining 9,212 patents.

As addressee, the first inventor listed on the patent document was chosen. To control for address changes we used the web version of the white pages as well as the EPOLINE® database of the EPO. The information was obtained using a questionnaire. Each inventor was provided with a cover letter together with an attached questionnaire. To date, 3,346 responses were received, resulting in a response rate of 32%. The sample contains 2,761 inventors who answered one questionnaire, 282 inventors with two questionnaires, 4 with three questionnaires, 1 inventor who answered four questionnaires, and 1 inventor who filled out 5 questionnaires.[7] Hence, the sample used in this paper is made-up of 3,049 different inventors (responsible for 3,346 EP patents).

The data from the questionnaire was merged with bibliographic and procedural information on the respective patents obtained from the online EPOLINE® database. The database

[7] Inventors who were responsible for more than one patent in the underlying time period and who were chosen more than once by stratified random sample, were provided with up to five questionnaires.

contains information on all published EP patent applications as well as all published PCT applications since the founding of the EPO in 1978. The dataset corresponds to the EPOLINE® data as of March 1st, 2003 and covers over 1,260,000 patent files with application dates ranging from June 1st, 1978 to July 25th, 2002. Inventor address data has been available up to 1999.

To trace the productivity of each inventor over time, the EPOLINE® database was further used to search for all patents belonging to the 3,049 inventors with priority dates between 1977 and 1999. The search procedure[8] resulted in a total of 29,971 EP patents.[9]

4.2 Age and Patent Propensity

In a number of studies, the relationship between age and productivity among technical personnel or scientists has been analyzed. Early findings show a maximum of productivity at the age of about 40 and a decline afterwards. This decline was explained by a decrease of motivation and risk-taking as well as by difficulties in keeping up with technological change (Dalton/Thompson 1971; Lehman 1966; Oberg, 1960). Later studies detected a curve with two modes, one before the age of 40, the second approximately at the age of 50 (Pelz/Andrews 1966; Vincent/Mirakhor 1972). These findings were criticized by Zuckerman and Merton (1972). Studying Nobel Prize winners, the authors showed that these scientists remained highly productive over time. A decline in productivity due to seniority was explained by differences between two groups: a small group of key scientists who increases or, at least, maintains their productivity level and another larger group showing a decrease in productivity over time. These findings are concordant with the Matthew Effect described above (Goldberg/Shenhav 1984; Zuckerman/Merton 1972).

Jones (2005) also uses data on Nobel Prize winners. Additionally, 20th century great inventors are included in the analysis. Results show an upward trend in the age at which scientists and engineers begin their careers. A reason for this delayed start is an increase of the age at the time of the highest educational degree. Thus, scientists and engineers spend more time on education. Indeed, this time shift is not compensated by a shift in the productivity of innovators beyond middle age. This leads to a decline of the overall innovative output of younger innovators. In particular, a 30% decline in life-cycle output over the 20th century is observed. Furthermore, the author finds that “the mean age of great achievement for both Nobel Prize winners and great technological inventors rose about 6 years over the course of the 20th century” (Jones 2005: 2).

---

8 The search procedure is described at length in Annex 1.

9 In 2003, Blind et al. conducted a study on behalf of the German Bundesministerium für Bildung und Forschung (BMBF), dealing with the apparent discrepancy between a rather modest increase in German R&D expenditure and the strong increase of patent applications filed by German firms. Within the scope of this analysis the researchers analyzed whether including patent data from the German Patent and Trademark Office reduces firm size effects. Results show an even larger concentration with respect to patent applications of large firms, compared to the EPO data. It is assumed, that increasing costs of an EP or PCT application prevents large firms from filing a patent for each and every invention, which does not apply to cheaper national patent applications (Blind et al. 2003). Although, using, EP patents could result in an underestimated productivity of inventors employed with small firms, including data from the Germen Patent and Trademark Office (GPTO) would even increase possible biases. Hence this paper refrains from using GPTO patent data.

Due to these controversial results[10], the age effect is studied more closely by using a sub-sample of the dataset. The sub-sample contains 579 inventors aged between 55 and 57 in 1999. Since the dataset comprises priority dates between 1977 and 1999, this group of inventors was especially appropriate to analyze age effects. Inventors aged between 55 and 57 in 1999, were between 33 and 35 years old in 1977. This allows for comparing productivity during an early stage of inventive lifetime to later stages. The total number of patents per inventor was divided into five groups with respect to the age of the inventor at the time of the priority date[11]. Five age groups were created: "30 to 35 years", "36 to 40 years", "41 to 45 years", "46 to 50 years", and "51 to 55 years". For an inventor aged 55 in 1999 the group "51 to 55 years", e.g., contains the number of patents of this inventor with priority year between 1995 and 1999.

Figure 2 contains two curves: "whole" represents whole patent counts, assigning the whole patent to each inventor of an inventor team. "Fractional", on the other hand, refers to fractional counts with respect to the number of co-inventors. For instance, if there were two inventors of a patent, each inventor is assigned half a patent. Fractional counts are said to be a more appropriate productivity measure, although this raises the problem of exactly quantifying the contributions of each co-inventor (Ernst et al. 2000). Both, whole and fractional counts will be included in the multivariate analysis.

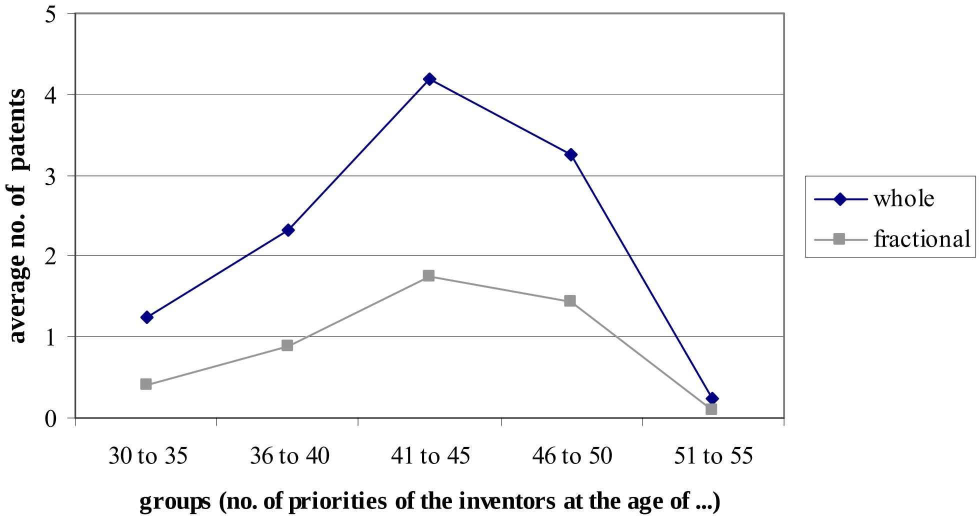

Figure 2: Average number of patents with respect to age (inventors 55 to 57 years old in 1999), N = 579

For this sub-sample, both curves show that the average number of patents of the inventors has a maximum at the age of 41 to 45, and declines after this point[12]. This result seems to confirm former studies which account the decline at the age of 40 to 45 to "difficulties in keeping up with the new knowledge generated by rapid technological change" or to "various psychological alterations with age", such as a reduction of motivation (Goldberg/Shenhav 1984: 111).

---

[10] See Goldberg/Shenhav (1984) for a summary of the relevant literature on the relationship between age and productivity.

[11] In case of more than one priority date, the earliest date was used to minimize the time span between the date of invention and the year included in the analysis.

[12] Part of the strong decrease in the last group (51 to 55 years) arises due to truncation of the data.

Another Issue that should be considered when measuring productivity is a changing patent propensity over time and across different industries. Due to strategic reasons, companies apply for more patents per Euro R&D expenditure. Hall (2004) uses U.S. patent data of about 1,400 U.S. manufacturing firms between 1980 and 1989 to explore the sources of patent growth in the U.S. since 1984. Results show that the increase of patent applications has taken place especially in the electrical, electronics, computing and scientific instruments industry. This “patent explosion” is assumed to be a result of a strategic shift in patenting behavior of U.S. firms in these industries (Hall 2004).

The EPO data also show an increase of the total number of priorities per year. Figure 3 displays a patent propensity index calculated for 30 technological areas between 1977 and 1999. The patent propensity index is defined as follows:

$$PP_{it} = \frac{number\ of\ \ prios_{it}}{number\ of\ \ prios_{i\,1985}}$$

where i is the technical area and t the priority year of the patents. $number\ of\ \ prios_{it}$ refers to the number of priorities in technical area i in year t. 1985 is used as a reference year[13], which means that the patent propensity index is 1 for all technical areas in 1985.

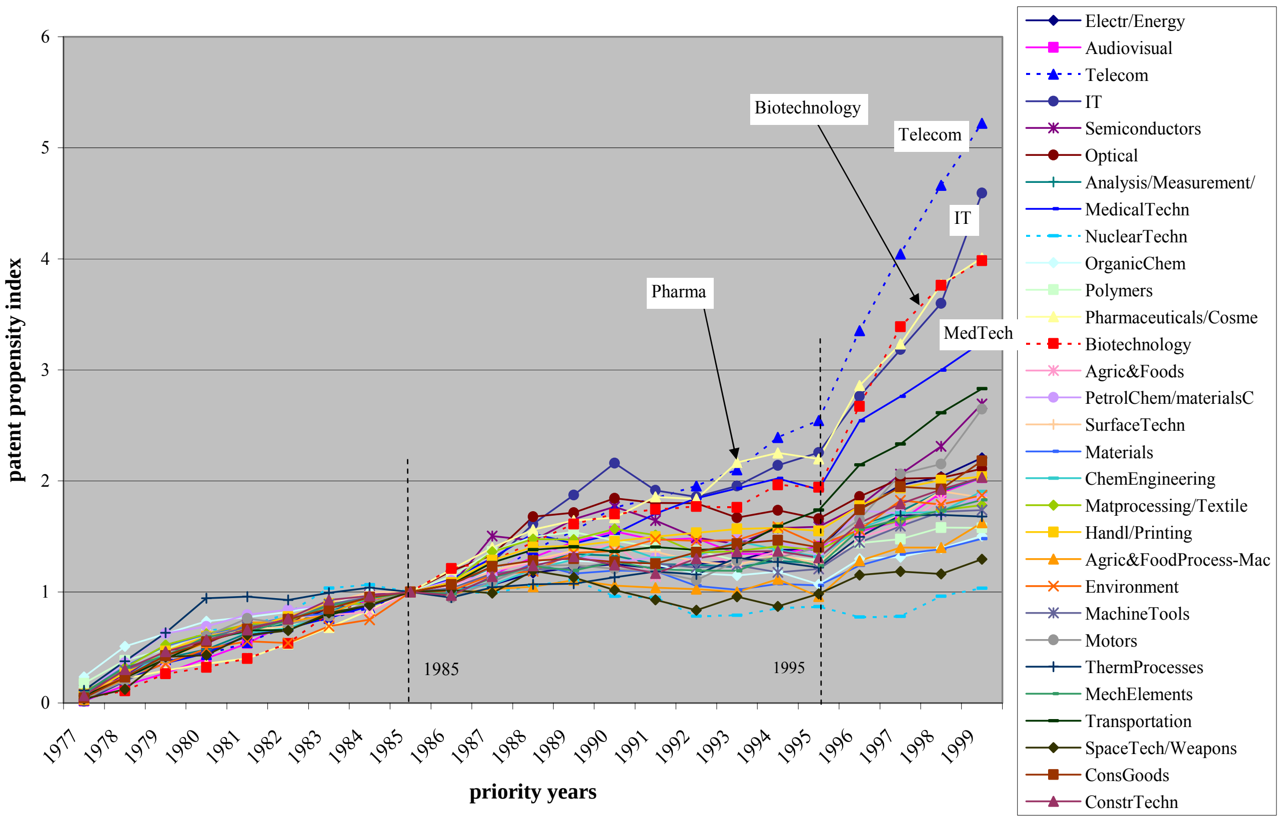

Figure 3: Patent propensity index per technological area (N = 1,186,321 EP patents)

---

[13] 1985 is used as a reference year, since the years between 1977 and 1984 are characterized by a strong increase in the number of EP priorities/applications, which is not caused by an increasing patent propensity but by the diffusion of the EP patent.

Results show that patent propensity increased mainly in the following technical areas: telecom, IT, biotechnology, pharmaceuticals, and medical technology. This finding is in accordance with the above described results of Hall (2004). Furthermore, Figure 3 shows that especially as from 1995, an increase becomes apparent in every technical area.

An increasing patent propensity over time leads to biased results, since younger inventors tend to patent more inventions than older ones when they were at the same age (Hall et al. 2005). To avoid these biases, it is necessary to control for patent propensity with respect to time and technical area. In the following analysis, three different productivity measures are used. First, the conventional measure, utilized in the literature, is employed, i.e. the number of patents per inventor is related to the age of the inventor. Second, a productivity index is calculated, which is based on the patent propensity index described above. Finally, quality adjusted patent counts are employed, weighting each patent with the number of citations it received from subsequent patents. Of course, the second and the third productivity measure will also be related to the age of the inventor.

4.3 Variables

Dependent Variables

1. **PRODUCTIVITY$_1$** (pure patent counts) **–** The variable is defined as the number of patents per inventor, divided by age minus 25. A way of justifying this measure would be the assumption that inventors become active at the age of 25.

$$PROD_1 = \frac{number\ of\ \ patents}{age_{1999} - 25} \qquad \text{(P.1)}$$

2. **PRODUCTIVITY$_2$** (patent propensity adjusted patent counts) **–** The productivity index is defined as the inverse ratio of the patent propensity index, added up for the total number of patents per inventor in the respective year and technical field. Subsequently, the adjusted number of patents is divided by age of the inventor in 1999 minus 25.

$$PROD_{it} = \frac{adjusted\ number\ of\ \ patents_{it}}{age_{1999} - 25}$$
$$= \frac{\sum_{i,t}\left(\left(\frac{number\ of\ \ prios_{it}}{number\ of\ \ prios_{i\,1985}}\right)^{-1} \cdot\ number\ of\ \ patents_{it}\right)}{age_{1999} - 25} \qquad \text{(P.2)}$$

where *i* refers to the technical area, and *t* to the priority year.

3. **PRODUCTIVITY$_3$** (citation counts) **–** Quality adjusted patent counts are defined as the number of citations[14] a patent received, added up for the total number of patents per inventor, divided by age of the inventor in 1999 minus 25. In accordance with Price (1976), who counts the publication of a paper as its first citation "success", the application of an EP patent is supposed to be its first patent citation. Patent counts are, therefore, weighted by the number of citations received plus one. This adjustment is necessary in order to calculate the logarithm of this variable.

$$PROD_3 = \frac{number\ of\ citations + 1}{age_{1999} - 25} \qquad \text{(P.3)}$$

Explanatory Variables

LEVEL OF EDUCATION - The questionnaire included a question asking the respondents for their highest degree. In order to simplify the analysis, the education variable was aggregated to three groups: (1) secondary school, high school diploma, or vocational training (reference group), (2) vocational academy or university studies, and (3) doctoral or postdoctoral studies.

INTER-FIRM MOBILITY – change of the employer. The respondents were further asked to indicate the number of times they changed their employer. A dummy variable was generated indicating whether the inventor changed the employer or not.[15]

CROSS-COUNTRY MOBILITY – temporary stay abroad. Respondents were asked to indicate whether they spent some time in a non-German speaking country. Again, a dummy variable was generated taking the value 1 for inventors who temporarily stayed abroad and 0 otherwise.

SOURCES OF KNOWLEDGE – university research, scientific literature, patent literature, customers, and competitors. The questionnaire included a question relating to the importance of different sources of knowledge for the development of an invention.[16] Answers were collected on a scale from one (absolutely not important) to five (very important). A dummy variable was created for each source of knowledge, combining categories 1 (absolutely not important) to 3 (partly important) as well as categories 4 (important) and 5 (very important). The latter implies a use of the respective knowledge source.

---

[14] In 1955 Eugene Garfield proposed the Science Citation Index providing access to a large number of scientific and technical journals (Narin 2000). Using the Science Citation Index, it became possible not only to count the number of articles published but also to include the number of citations a published article receives from subsequent articles. Since then numerous researchers used citation data in order to account for qualitative productivity (Turner/Mairesse 2002, Allison/Stewart 1974, Bayer/Folger 1966, Sher/Garfield 1966). This method is transferred to patent counts or inventive productivity.

[15] An inter-firm mobility variable, containing the ordinal information (the inventor did not change employer, changed the employer once, twice, three times or more), was also tested but did not lead to different results.

[16] Although the answers of the questionnaire were related to specific patents, the answers seem to be transferable to all patents of an inventor. It is assumed that inventors basically tend to use special sources of knowledge, for example, due to positive experiences in the past. This assumption proves true, when comparing the answers of inventors who filled out more than one (five at the most) questionnaires. The different sources of knowledge are found to be equally important for all surveyed patents per inventor. Those answers that do not show a perfect match are at least highly correlated.

FIRM SIZE - number of employees. The firm size was also obtained from the questionnaire. A set of eight dummy variables was generated in order to account for variation across different firm sizes. The intervals range from “less than 50 employees” to “more than 50,000 employees”. Except for the first group (“less than 50 employees” = reference group), the dummies were included in the analysis.

OPPOSITIONS - The variable contains the share of patents per inventor that were opposed by a third party within the opposition term of nine months after grant.

STATUS - This variable provides information on the status of the patent applications. Two share variables were included accounting for the share of patents that were either refused by the examiner or withdrawn by the applicant, e.g., due to the results of the search report. The status variables as well as the opposition variable are included to control for the value of the patents.

5 Descriptive Statistics und Multivariate Results

5.1 Descriptive Statistics

The empirical analysis is based on the responses of 2,630[17] inventors who are responsible for a total of 26,601 EP patents. Table 1 presents selected descriptive statistics of the variables described in the previous section. Each inventor is on average responsible for 10 EP patents.[18] The total number of patents per inventor ranges between 1 and 258. The standard deviation amounts to 16.3. Fractional patent counts range between 0.1 and 80.4 and have its mean at 4.4. The adjusted patent counts, controlling for the increasing number of patent applications over time and across 30 technical fields, lies between 0.3 and 223.7 (mean = 8.2, s.d. = 15.3). The patents per inventor received an average of 13.7 citations, ranging from 0 to 721. Additional information is provided on the legal status of the patents. On average 7% of the inventors’ patents were opposed by a third party (s.d. = 0.2), on average 11% (s.d. = 0.2) of the patents had been withdrawn, and 2% (s.d. = 0.1) had been refused by the EPO. Statistics of the EPO reveal that on average 29.7% of EP patent applications between 1980 and 1990 had been withdrawn by the applicant and 5.2% had been refused by the EPO (Harhoff/Wagner 2003). A possible reason for the low rates of withdrawal and refusal within this data may be the oversampling of important patents.

17 2,630 of the 3,349 questionnaires were filled out completely with regard to the above described variables.

18 In the PatVal questionnaire the inventor was asked to provide the number of EP patents he is responsible for. The correlation between the number of patents deriving from the EPOLINE® database and the number of EP patents from the questionnaire reveals a correlation coefficient of 0.74 (0.76 when the logarithm of the variables is used). The high correlation coefficient points at a high match between the information from the two different data sources.

Variable	Mean	S.D.	Min.	Max.
number of patents (whole)	10.14	16.29	1	258
number of patents (fractional)	4.36	6.04	0.09	80.43
adjusted number of patents	8.24	15.32	0.30	223.68
number of citations received (whole)	13.66	32.80	0	721
number of citations received (fractional)	5.17	9.93	0	135.01
age of the inventor in 1999	49.78	9.74	28	80
share of patents opposed	0.07	0.17	0	1
share of patents refused	0.02	0.05	0	0.50
share of patents withdrawn	0.11	0.16	0	0.75
secundary school/vocational training / high school diploma	0.13		0	1
university studies	0.53		0	1
doctoral/post-doctoral studies	0.34		0	1
inter-firm mobility	0.17		0	1
cross-country mobility	0.26		0	1
source of knowledge - universities	0.23		0	1
source of knowledge - literature	0.62		0	1
source of knowledge - other patents	0.65		0	1
source of knowledge - customers	0.73		0	1
source of knowledge - competitors	0.56		0	1
firm size (no. of employees)	47,776	93,649	1	550,000

Table 1: Descriptive statistics (N = 2,630)[19]

The responding inventors are aged between 28 and 80 in 1999 with a mean at 49.8 years (s.d. = 9.7). Furthermore, the respondents are characterized by a very high level of education. Asked about their highest level of education, 13% of the inventors indicated “high school diploma” or “vocational training”. More than 50% have a university degree and 34% a doctoral or post-doctoral degree. The inter-firm mobility dummy reveals that 17% of the respondents changed their employer, 26% temporarily stayed abroad. Users of the patented inventions as well as other patent documents turned out to be the most popular sources of knowledge. 73% of the inventors believe customers or users to be an important source of knowledge and 65% make use of other patent documents. Only 23% of the respondents believe university research to be important for making inventions. On average the patent assignees’ firms have 47,776 employees. The number of employees ranges between 1 and 550,000. The high standard deviation, amounting to 93,649, can be interpreted as a sign of a strongly heterogeneous distribution of this variable. Therefore, for the following multivariate analysis firm size groups are used.

[19] 2,630 inventors filled out the questionnaire completely with respect to the included variables.

	number of patents	number of citations	adjusted number of patents	age in 1999	share_ opposition	share_ refused	share_ withdrawn	firm size
number of patents	1.000							
number of citations	0.889*	1.000						
adjusted number of patents	0.969*	0.881*	1.000					
age in 1999	0.131*	0.136*	0.178*	1.000				
share_opposition	-0.058*	-0.016	-0.043*	0.064*	1.000			
share_refused	0.081*	0.059*	0.070*	0.059*	-0.032	1.000		
share_withdrawn	0.188*	0.134*	0.171*	0.078*	-0.097*	0.061*	1.000	
firm size	0.066*	0.048*	0.065*	-0.061*	-0.059*	0.049*	0.001	1.000

* significant at 5% or lower

Table 2: Pearson correlation coefficients (N = 2,630)

Table 2 lists the Pearson correlation coefficients for interval scaled variables. The dependent variables "number of patents" and "adjusted number of patents" are positively correlated with age, the variables concerning the legal status of the EP patents, and firm size. The share of patents opposed by third parties possesses a negative correlation. Regarding "number of citations", Table 2 presents quite similar results. The number of citations is also negatively correlated with the share of patents opposed, although not significantly. Overall, the correlation coefficients of the independent variables are quite small, indicating variables with orthogonal information.

The following two figures aim at a more detailed description of the dependent variables.

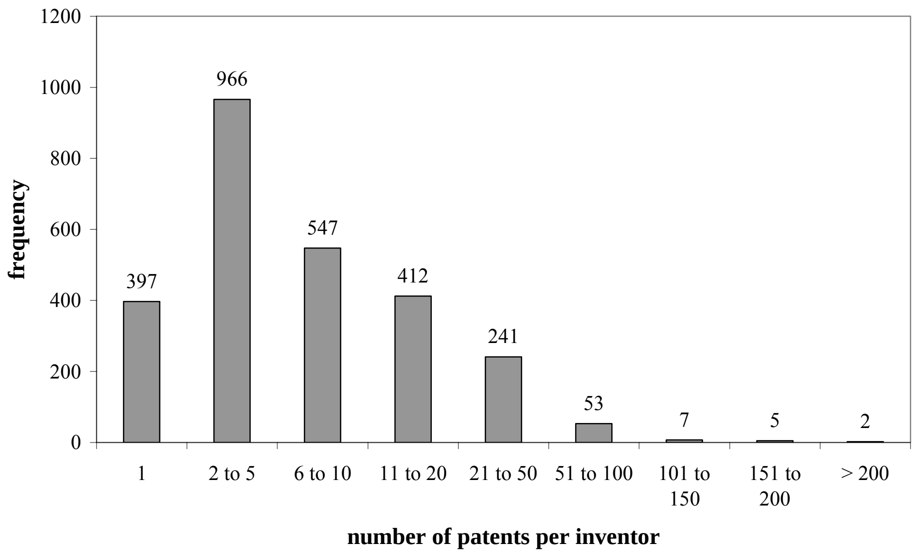

Figure 4: Distribution of whole patent counts (N = 2,630)

Figure 4 reports the distribution of the number of patents per inventor. Due to the requirements of the stratified random sample, we used for the PatVal project, each inventor is (jointly) responsible for at least one patent with priority date between 1993 and 1997. Assigning the whole patent to each inventor of an inventor team, 15% of the inventors are responsible for 1 patented invention. The bigger part (58%) holds 2 to 10 patents. Only 3% are responsible for more than 50 patents.

Figure 5 displays a histogram of the number of citations received by the total number of patents of each inventor. The remaining tail (more than 60 forward citations) is displayed separately in the right upper corner. Nearly 20% of the inventors hold patents that received no citations at all. Only 2.1% of the inventors are responsible for patents that received more than 100 citations.

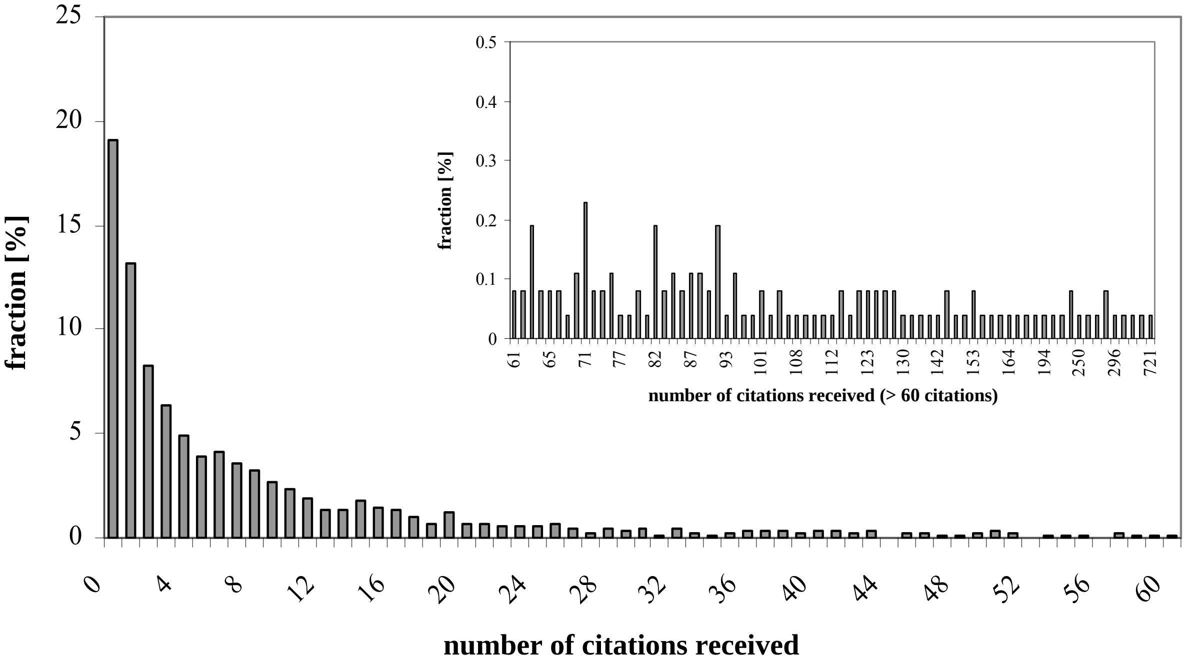

Figure 5: Distribution of citations received by the total number of patents per inventor (N = 2,630)

Tables 3 and 4, which are described in the following, give first insights into relationships between dependent and independent variables. Table 3 displays the relationship between the level of education of the inventors and whole patent counts. Whereas inventors who earned a high school diploma or served a vocational training on average have 6 patents, inventors with a university degree are slightly more productive (mean = 7 patents). Inventors with doctoral or postdoctoral studies are more than twice as productive compared to the two other groups, holding an average of 16 patents (ANOVA: F = 100.17, p = 0.000).

Level of education (groups)	**Number of patents per inventor (whole counts)**	
	Number of observations	**Mean**
secondary school/vocational training / high school diploma	333	6.09
university studies	1,396	7.24
doctoral/post-doctoral studies	901	16.12
Total	2,630	10.14

Note: In an ANOVA, the effect of the level of education turned out to be highly significant (F = 100.17, p = 0.000).

Table 3: Number of patents (whole counts) by level of education (N = 2,630)

Table 4 presents the relationship between inventive output (number of patents) and firm size of the inventors' employer (number of employees). Table 2 contained a first reference to a positive correlation between the dependent variable and firm size. Results show that inventive output increases almost monotonically with an increasing firm size (ANOVA: $F = 22.42$, $p = 0.000$). This result can be considered as a first indicator in favor of hypothesis 5.

Firm Size in number of employees	**Number of patents per inventor (whole counts)**	
	Number of observations	**Mean**
Less than 50 employees	152	4.95
51 – 250 employees	231	4.96
251 – 500 employees	196	6.58
501 – 1,500 employees	361	6.65
1,501 – 5,000 employees	422	8.11
5,001 – 10,000 employees	231	11.07
10,001 – 50,000 employees	488	14.74
More than 50,000 employees	549	14.39
Total	2,630	10.14

Note: In an ANOVA, the effect of the firm size turned out to be highly significant ($F = 22.42$, $p = 0.000$).

Table 4: Number of patents (whole counts) by firm size (N = 2,630)

5.2 Multivariate Specification

Following the findings of Lotka (1926) and Narin and Breitzman (1995), a log-normal distribution of productivity is assumed. The hypotheses derived in section 3 lead to the following specification:

$$\log(PROD_{m\,jkt}) = \beta_0 + \beta_1 * \log(\text{age}_{j\,1999}\text{-}25) + \beta_2 * \text{share\_oppo}_t + \beta_3 * \text{share\_refused}_t$$
$$+ \beta_4 * \text{share\_withdrawn}_t + \beta_5 * \text{d\_educ}_{jt} + \beta_6 * \text{d\_mobility}_{jt} + \beta_7 * \text{d\_source\_knowledge}_{jt}$$
$$+ \beta_8 * \text{d\_firmsize}_{kt} + \beta_9 * \text{share\_tech\_area}_t + \varepsilon_{jkt}$$

referring to the productivity of an inventor *j* employed with company *k* in the year *t*. For $m = 1$ patent productivity (not adjusted) (P.1) is used as a dependent variable. $m = 2$ points at a second model, employing the productivity index (P.2). The citation productivity (P.3) is used for $m = 3$. The above described log-specification is estimated using an OLS Regression.

Table 5 presents the results of Model (1) which contains the independent variables required for testing the hypotheses. Model (2), displayed in Table 6, additionally includes variables controlling for variation between technological areas.

The main differences between the results of Model (1) and (2) arise with respect to the level of education. University studies do not significantly influence productivity. In case the productivity index (P2) is used as a dependent variable, the coefficient becomes significant after including the control variables. Additionally, the effect of university studies on productivity increases irrespective of the dependent variable. Doctoral studies, on the contrary, become less important after including the control variables. Furthermore, the coefficients of firm size decrease. Both effects may arise due to large pharmaceutical and chemical enterprises, characterized by a large number of employees and an above average fraction of researchers who have a doctoral degree. In the following, the results of Model (2) are described.

Model (2a) and Model (2b) were included to display differences arising due to the use of the productivity index (P.2) as a dependent variable. The index was constructed to control for a different patent propensity over time and between different technical areas. Comparing the Models (2a) and (2b) with the remaining models reveals that the coefficient of (age-25) is -0.6 with respect to pure patent counts (-0.5 with respect to fractional patent counts) and -0.4 (-0.3) in case the index is employed. Age was used as a reference time in order to compare (adjusted) patent counts between inventors. Results indicate that the coefficient of age is biased, when not controlling for patent propensity changes. Therefore, the specifications using the productivity index as a dependent variable are preferred over those using pure patent productivity.

Models (2c) and (2d) refer to the dependent variable logarithmic productivity index (P2), whereas the Models (2e) and (2f) refer to the logarithmic citation productivity (P3). Besides whole counts, fractional counts were included. At first, the results concerning the productivity index are reported. *log(age-25)*[20] was included in the regression to estimate a coefficient for age instead of assuming the coefficient to be 1, i.e., to take a directly proportional relationship between adjusted patent counts and age for granted. A coefficient of -0.4 implies that the number of adjusted patents rises less than directly proportional with age (slope: 1-0.4 = 0.6). Thus, when age increases by 1%, the adjusted number of patents rises by 0.6%. The effect is significant at the 1% level.

Whereas a marginal increase of the share of opposed patents decreases productivity by 0.36%, the share of patents that were either refused by the EPO or withdrawn by the applicant increases the productivity by 2.96% and 1.68%, respectively.

A set of dummies for level of education (university studies, (post-) doctoral studies, reference group: high school diploma or less) shows that inventors who earned a university degree are 16% more productive than the inventors of the control group. Inventors who did doctoral or post-doctoral studies are 34% more productive, again compared to the control group. Using fractional patent counts cuts the effect of doctoral studies down to 26%. The coefficients are significant at the 1% level. These findings fully support hypothesis 1, stating that inventors with a higher level of education tend to show a higher productivity.

[20] In the following referred to as age.

Inventors who are mobile between firms or countries show a lower productivity. Whereas the inter-firm mobility coefficient is not significant, cross-country mobility significantly decreases productivity by 8%. Since the effect is only significant at the 10% level and the mobility coefficients of the remaining models are consistently insignificant, it is assumed that hypotheses 2 and 3 are not supported. Testing inter-firm mobility and cross-country mobility jointly using a Wald test confirms insignificance ($\chi^2 = 1.94$, $P = 0.144$).

As expected, exploiting the knowledge from other patent documents and competitors' knowledge increases productivity. Inventors who make use of patent documents are 23% more productive than inventors who do not use this source of knowledge. Inventors attaching great importance to competitors as an external source of knowledge are 12% more productive compared to inventors who do not use information derived from competitors. Both coefficients are significant at the 1% level. After using the control variables, including users in the invention process does not have a significant effect on productivity. Only Model (1b) provides a slight evidence for a negative effect. Making use of scientific literature decreases productivity by 9% (whole counts) or 15% (fractional counts).
Whereas Model (1c) supports the proposition that applying scientific knowledge requires absorptive capacity, the coefficients of the interaction terms become insignificant after including the control variables. Doctoral or post-doctoral studies increase the influence of scientific literature on productivity by 58.7%. The interaction between doctoral studies and spillovers from university research is not significant at the 10% level. These results, at least in part, support hypotheses 4a and 4b.

A set of dummies for FIRMSIZE, measured by number of employees (51 – 250 employees, 251 – 500 employees, 501 – 1,500 employees, 1,501 – 5,000 employees, 5,001 – 10,000 employees, 10,001 – 50,000 employees, more than 50,000 employees), show that an increasing firm size monotonically rises productivity, compared to the reference group (less than 50 employees). The coefficients (except for 51 – 250 employees) are significant at the 1% level. Interestingly, the firm size effects in part become insignificant, when fractional patent counts are employed (Model 2d). Those coefficients that are still significant have a value which is a third compared to the results using whole patent counts. For instance, whereas inventors employed with firms that have 5,001 to 10,000 employees are 50.5% more productive compared to the control group, inventors are only 18.1% more productive, when fractional counts are considered. Overall, hypothesis 5 is also supported by the data.

The control variables, accounting for the share of patents the inventors applied for in each technical area, test for heterogeneity that would otherwise lead to biased results of the explanatory variables. Table 7 displays that especially the share of patents in organic chemistry, medical technology, as well as environment do have a significant effect on productivity. Although not interpretable, it is important to mention that the coefficients do not show contra-intuitive signs.

	Model (1)					
	(a)	(b)	(c)	(d)	(e)	(f)
	log(productivity) [FULL]	**log(productivity)** [FRACTIONAL]	**log(productivity index)** [FULL]	**log(productivity index)** [FRACTIONAL]	**log(citation productivity)** [FULL]	**log(citation productivity)** [FRACTIONAL]
productivity age (log(age-25))	-0.629***	-0.531***	-0.445***	-0.347***	-0.492***	-0.407***
	[0.042]	[0.045]	[0.045]	[0.047]	[0.051]	[0.052]
share of patents opposed	-0.453***	-0.454***	-0.378***	-0.384***	-0.145	-0.141
	[0.095]	[0.109]	[0.104]	[0.117]	[0.135]	[0.140]
share of patents refused	2.645***	2.836***	2.725***	2.914***	3.160***	3.433***
	[0.495]	[0.461]	[0.527]	[0.482]	[0.592]	[0.534]
share of patents withdrawn	1.754***	1.652***	1.861***	1.749***	1.864***	1.746***
	[0.126]	[0.128]	[0.135]	[0.137]	[0.149]	[0.149]
level of education (reference group: high school diploma or less)						
university studies	0.114**	0.106*	0.096	0.091	0.114*	0.105
	[0.056]	[0.061]	[0.059]	[0.063]	[0.069]	[0.071]
doctoral/postdoctoral studies	0.450***	0.283***	0.458***	0.290***	0.598***	0.436***
	[0.092]	[0.096]	[0.097]	[0.100]	[0.109]	[0.110]
inter-firm mobility	-0.055	-0.029	-0.074	-0.047	-0.018	0.001
	[0.050]	[0.053]	[0.053]	[0.055]	[0.061]	[0.062]
cross-country mobility	-0.059	-0.063	-0.055	-0.059	-0.035	-0.048
	[0.044]	[0.046]	[0.046]	[0.048]	[0.053]	[0.054]
source of knowledge - universities	-0.095	-0.091	-0.067	-0.064	-0.126*	-0.120
	[0.062]	[0.067]	[0.065]	[0.070]	[0.075]	[0.079]
doctoral studies * knowledge_university	-0.011	-0.004	-0.083	-0.077	-0.015	-0.006
	[0.095]	[0.100]	[0.101]	[0.105]	[0.115]	[0.117]
source of knowledge - literature	-0.088*	-0.164***	-0.073	-0.154***	-0.068	-0.149**
	[0.048]	[0.052]	[0.050]	[0.054]	[0.059]	[0.061]
doctoral studies * knowledge_literature	0.172*	0.098	0.202**	0.136	0.164	0.079
	[0.095]	[0.098]	[0.100]	[0.103]	[0.113]	[0.113]
source of knowledge - other patents	0.268***	0.199***	0.291***	0.230***	0.334***	0.266***
	[0.043]	[0.046]	[0.045]	[0.048]	[0.052]	[0.054]
source of knowledge - user	-0.047	-0.078*	-0.034	-0.065	-0.062	-0.085
	[0.044]	[0.046]	[0.047]	[0.048]	[0.054]	[0.055]
source of knowledge - competitors	0.127***	0.139***	0.119***	0.130***	0.139***	0.149***
	[0.042]	[0.045]	[0.044]	[0.046]	[0.050]	[0.052]
firm size in number of employees (reference group: less than 51 employees)						
51 - 250 employees	0.097	-0.011	0.090	-0.024	0.028	-0.086
	[0.091]	[0.098]	[0.094]	[0.101]	[0.113]	[0.117]
251 - 500 employees	0.337***	0.146	0.363***	0.164	0.321***	0.128
	[0.097]	[0.101]	[0.102]	[0.105]	[0.123]	[0.124]
501 - 1,500 employees	0.370***	0.031	0.409***	0.070	0.396***	0.049
	[0.086]	[0.089]	[0.090]	[0.093]	[0.109]	[0.109]
1,501 - 5,000 employees	0.470***	0.107	0.489***	0.127	0.555***	0.196*
	[0.084]	[0.087]	[0.088]	[0.092]	[0.106]	[0.106]
5,001 - 10,000 employees	0.585***	0.230**	0.573***	0.215**	0.685***	0.322***
	[0.100]	[0.104]	[0.104]	[0.107]	[0.125]	[0.125]
10,001 - 50,000 employees	0.716***	0.209**	0.747***	0.245***	0.844***	0.334***
	[0.087]	[0.087]	[0.091]	[0.091]	[0.109]	[0.106]
more than 50,000 employees	0.690***	0.249***	0.720***	0.281***	0.825***	0.374***
	[0.086]	[0.087]	[0.090]	[0.091]	[0.108]	[0.106]
Constant	-0.557***	-1.112***	-1.510***	-2.067***	-0.603***	-1.124***
	[0.165]	[0.173]	[0.174]	[0.181]	[0.199]	[0.201]
Observations	2630	2630	2630	2630	2630	2630
R-squared	0.312	0.184	0.287	0.166	0.284	0.175
F-test (df)	56.69(22)	27.65(22)	49.26(22)	24.28(22)	47.38(22)	25.38(22)

Robust standard errors in brackets
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 5: OLS Regression (Model 1 with heteroskedasticity-robust standard errors) (N = 2,630)

	Model (2)					
	(a)	(b)	(c)	(d)	(e)	(f)
	log(productivity) [FULL]	**log(productivity)** [FRACTIONAL]	**log(productivity index)** [FULL]	**log(productivity index)** [FRACTIONAL]	**log(citation productivity)** [FULL]	**log(citation productivity)** [FRACTIONAL]
productivity age (log(age-25))	-0.609***	-0.512***	-0.445***	-0.347***	-0.463***	-0.379***
	[0.042]	[0.045]	[0.044]	[0.047]	[0.050]	[0.051]
share of patents opposed	-0.417***	-0.427***	-0.359***	-0.374***	-0.094	-0.100
	[0.096]	[0.111]	[0.104]	[0.118]	[0.136]	[0.142]
share of patents refused	2.806***	2.903***	2.962***	3.056***	3.328***	3.504***
	[0.498]	[0.461]	[0.534]	[0.492]	[0.591]	[0.536]
share of patents withdrawn	1.613***	1.605***	1.675***	1.657***	1.678***	1.656***
	[0.124]	[0.130]	[0.131]	[0.137]	[0.148]	[0.151]
level of education (reference group: high school diploma or less)						
university studies	0.166***	0.134**	0.161***	0.135**	0.177***	0.144**
	[0.057]	[0.061]	[0.059]	[0.063]	[0.068]	[0.071]
doctoral/postdoctoral studies	0.361***	0.282***	0.338***	0.258***	0.482***	0.407***
	[0.093]	[0.097]	[0.095]	[0.100]	[0.110]	[0.112]
inter-firm mobility	-0.035	-0.036	-0.041	-0.040	0.005	-0.001
	[0.050]	[0.053]	[0.051]	[0.055]	[0.060]	[0.062]
cross-country mobility	-0.076*	-0.065	-0.080*	-0.070	-0.058	-0.057
	[0.043]	[0.046]	[0.045]	[0.048]	[0.052]	[0.053]
source of knowledge - universities	-0.075	-0.080	-0.057	-0.063	-0.097	-0.100
	[0.062]	[0.067]	[0.065]	[0.070]	[0.074]	[0.079]
doctoral studies * knowledge_university	0.042	0.018	0.005	-0.020	0.053	0.031
	[0.093]	[0.100]	[0.097]	[0.104]	[0.112]	[0.116]
source of knowledge - literature	-0.105**	-0.156***	-0.093*	-0.149***	-0.099*	-0.155**
	[0.048]	[0.052]	[0.050]	[0.054]	[0.059]	[0.061]
doctoral studies * knowledge_literature	0.081	0.053	0.102	0.081	0.052	0.015
	[0.093]	[0.099]	[0.096]	[0.101]	[0.110]	[0.112]
source of knowledge - other patents	0.216***	0.186***	0.228***	0.205***	0.267***	0.241***
	[0.043]	[0.046]	[0.045]	[0.048]	[0.052]	[0.054]
source of knowledge - user	0.007	-0.046	0.013	-0.040	0.001	-0.046
	[0.045]	[0.048]	[0.046]	[0.049]	[0.054]	[0.056]
source of knowledge - competitors	0.129***	0.123***	0.120***	0.113**	0.139***	0.131**
	[0.041]	[0.044]	[0.043]	[0.046]	[0.050]	[0.052]
firm size in number of employees (reference group: less than 51 employees)						
51 - 250 employees	0.074	-0.018	0.052	-0.046	0.009	-0.089
	[0.091]	[0.097]	[0.094]	[0.100]	[0.113]	[0.117]
251 - 500 employees	0.309***	0.133	0.312***	0.129	0.278**	0.101
	[0.099]	[0.102]	[0.104]	[0.106]	[0.125]	[0.125]
501 - 1,500 employees	0.344***	0.014	0.363***	0.033	0.363***	0.025
	[0.087]	[0.089]	[0.092]	[0.093]	[0.111]	[0.110]
1,501 - 5,000 employees	0.439***	0.109	0.448***	0.120	0.514***	0.191*
	[0.086]	[0.087]	[0.090]	[0.092]	[0.108]	[0.107]
5,001 - 10,000 employees	0.512***	0.192*	0.505***	0.181*	0.596***	0.269**
	[0.102]	[0.104]	[0.107]	[0.108]	[0.128]	[0.126]
10,001 - 50,000 employees	0.588***	0.173*	0.595***	0.185**	0.677***	0.261**
	[0.090]	[0.089]	[0.094]	[0.093]	[0.112]	[0.109]
more than 50,000 employees	0.611***	0.217**	0.643***	0.251***	0.721***	0.318***
	[0.090]	[0.089]	[0.094]	[0.094]	[0.113]	[0.110]
	… to be continued …					

Robust standard errors in brackets
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 6: OLS Regression (Model 2 with heteroskedasticity-robust standard errors) (N = 2,630) (Part I)

	Model (2) …continued…					
	(a)	(b)	(c)	(d)	(e)	(f)
	log(productivity) [FULL]	**log(productivity)** [FRACTIONAL]	**log(productivity index)** [FULL]	**log(productivity index)** [FRACTIONAL]	**log(citation productivity)** [FULL]	**log(citation productivity)** [FRACTIONAL]
share of patents in the following technical area (reference group: therm. Processes)						
electricity/energy	0.026	-0.058	0.007	-0.074	0.110	0.024
	[0.140]	[0.151]	[0.143]	[0.154]	[0.166]	[0.171]
audiovisual	0.309	0.354	0.340	0.389*	0.455	0.489*
	[0.243]	[0.224]	[0.259]	[0.233]	[0.309]	[0.282]
telecom	0.179	0.353*	-0.339*	-0.157	0.389*	0.561**
	[0.191]	[0.199]	[0.196]	[0.206]	[0.215]	[0.223]
information technology	0.026	0.106	-0.462**	-0.388	0.148	0.248
	[0.220]	[0.262]	[0.227]	[0.271]	[0.279]	[0.320]
semiconductors	-0.131	-0.158	-0.313	-0.331	-0.151	-0.178
	[0.244]	[0.274]	[0.241]	[0.275]	[0.302]	[0.326]
optical	0.414*	0.344	0.213	0.152	0.708***	0.653***
	[0.229]	[0.223]	[0.240]	[0.236]	[0.256]	[0.249]
analysis/measurem./control technology	0.070	-0.052	0.094	-0.027	0.158	0.043
	[0.149]	[0.160]	[0.150]	[0.163]	[0.180]	[0.187]
medical technology	0.206	0.229	-0.158	-0.126	0.407**	0.427**
	[0.170]	[0.182]	[0.178]	[0.189]	[0.205]	[0.215]
nuclear technology	-0.396	-0.152	0.097	0.311	-0.423	-0.130
	[0.292]	[0.337]	[0.282]	[0.352]	[0.295]	[0.362]
organic chemistry	0.954***	0.389**	1.096***	0.537***	1.185***	0.602***
	[0.169]	[0.173]	[0.172]	[0.178]	[0.199]	[0.197]
polymers	0.532***	0.049	0.580***	0.099	0.845***	0.347*
	[0.157]	[0.160]	[0.160]	[0.164]	[0.186]	[0.183]
pharmaceuticals/cosmetics	0.433**	0.145	-0.015	-0.308	0.900***	0.600**
	[0.210]	[0.219]	[0.217]	[0.226]	[0.253]	[0.247]
biotechnology	0.140	-0.578	-0.218	-0.929**	0.409	-0.339
	[0.381]	[0.408]	[0.411]	[0.430]	[0.452]	[0.490]
agriculture/food	-0.425*	-0.500*	-0.404	-0.479*	-0.547*	-0.665**
	[0.257]	[0.281]	[0.266]	[0.285]	[0.327]	[0.309]
petrol/materials chemistry	0.650***	0.310	0.602***	0.271	0.951***	0.620***
	[0.191]	[0.196]	[0.197]	[0.203]	[0.236]	[0.234]
surface technology	0.235	-0.076	0.185	-0.116	0.213	-0.109
	[0.200]	[0.213]	[0.205]	[0.219]	[0.247]	[0.253]
materials	-0.153	-0.459***	0.019	-0.284	-0.129	-0.457**
	[0.159]	[0.169]	[0.163]	[0.173]	[0.196]	[0.198]
chem. engineering	0.139	-0.060	0.166	-0.039	0.422*	0.195
	[0.190]	[0.209]	[0.197]	[0.216]	[0.223]	[0.239]
material processing/textiles/paper	0.049	-0.082	-0.018	-0.143	0.290	0.154
	[0.150]	[0.157]	[0.153]	[0.160]	[0.180]	[0.182]
handling/printing	0.149	0.093	-0.018	-0.069	0.207	0.150
	[0.143]	[0.151]	[0.145]	[0.154]	[0.169]	[0.172]
agric/food process-machines	0.202	0.249	0.424**	0.474**	0.528**	0.549**
	[0.209]	[0.206]	[0.213]	[0.209]	[0.255]	[0.241]
environment	-0.346*	-0.465**	-0.485**	-0.589***	-0.202	-0.308
	[0.184]	[0.199]	[0.193]	[0.209]	[0.221]	[0.217]
machine tools	-0.055	-0.164	-0.001	-0.116	-0.088	-0.186
	[0.148]	[0.162]	[0.150]	[0.165]	[0.176]	[0.184]
motors	0.040	0.007	-0.022	-0.053	0.176	0.136
	[0.171]	[0.172]	[0.173]	[0.175]	[0.202]	[0.197]
mech. elements	0.043	0.001	0.046	0.006	0.185	0.135
	[0.149]	[0.162]	[0.150]	[0.164]	[0.175]	[0.181]
transportation	0.078	-0.109	-0.139	-0.323**	0.262	0.057
	[0.138]	[0.149]	[0.141]	[0.152]	[0.165]	[0.168]
space technology/weapons	0.015	-0.202	0.312	0.105	-0.045	-0.237
	[0.240]	[0.269]	[0.254]	[0.284]	[0.304]	[0.326]
consumer goods	0.107	0.076	0.058	0.032	0.196	0.169
	[0.163]	[0.168]	[0.166]	[0.173]	[0.192]	[0.193]
construction technology	0.243	0.330**	0.178	0.269	0.298	0.378**
	[0.160]	[0.167]	[0.164]	[0.170]	[0.188]	[0.191]
Constant	-0.710***	-1.192***	-1.510***	-1.999***	-0.897***	-1.339***
	[0.201]	[0.209]	[0.208]	[0.216]	[0.240]	[0.240]
Observations	2630	2630	2630	2630	2630	2630
R-squared	0.350	0.210	0.339	0.198	0.327	0.206
F-test (df)	28.94(51)	14.10(51)	27.07(51)	12.95(51)	25.24(51)	13.76(51)

Robust standard errors in brackets
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 7: OLS Regression (Model 2 with heteroskedasticity-robust standard errors) (N = 2,630) (Part II)

Using the logarithm of citation productivity (Models (2e) and (2f)), which accounts for the quality adjusted productivity of an inventor, the share of patents that were opposed by a third party is no longer significant. This is not surprising, since the incident of an opposition can be interpreted as a proxy for the value of a patent (Harhoff/Hall 2003). Citation counts are also a proxy for patent value. It is therefore assumed that the share of oppositions is not required as a control variable within this specification.

Doctoral studies do have a larger impact on productivity (compared to the control group), when citations counts are employed. In particular, the effect increases by more than one third compared to Model (2c). The increase persists when fractional citation counts are used. The mobility variables, when related to citation counts, are again not significant (Wald test: $\chi^2 = 0.63$, $P = 0.5338$).

Compared to Model (2c), firm size coefficients are minor with respect to small firms but rise with increasing firm size.

6 Conclusion

The purpose of this paper was to analyze the determinants of inventor productivity, including inventor related and environmental determinants. To do so, at first, a productivity index was calculated controlling for patent propensity changes over time and patent propensity differences between 30 different technical areas. Especially industries such as telecom, IT, biotechnology/pharmaceuticals, and semiconductors are characterized by a disproportionate increase in EP patent applications per year. The adjusted patent counts were subsequently related to the age of the inventor minus 25 (the age of 25 refers to the starting point of inventive activity). Citations were employed as a second productivity measure, dividing the total number of citations per inventor by age minus 25. Fractional patent counts turned out to be an appropriate means to adequately capture firm size effects.

Two different hypotheses proposed by Allison and Steward (1974) were applied to distinguish between environmental and inventor specific factors determining productivity differences. According to the *Sacred Spark Hypothesis,* differences in productive capacity arise due to substantial, predetermined dissimilarities. The *Accumulative Advantage Hypothesis* proposes that the allocation of recognition and resources make highly productive scientists even more productive or lead at least to maintenance of output productivity. Considering the results of this paper, both perceptions have explanatory power with respect to productivity differences between inventors:

The level of education, representing the Sacred Spark Hypothesis, has a strong influence on inventor productivity. Inventors who attended university or who have a doctoral degree hold more patents and receive more citations. Doctoral studies become less important after controlling for industry differences. It is assumed, that this outcome is caused by industry differences: The proportion of students who earned a doctoral or post-doctoral degree is highest in chemistry, pharmaceuticals and medicine. At the same time, the chemical and pharmaceutical industry produces the highest patent rates as well as the most important patents.

Although not significantly (except for Model (1b)), including users in the innovation process tends to result in a decreasing productivity. Initially, this result was not expected since the involvement of users in the R&D process leads to knowledge spillovers. According to Foray

(2004), user innovations are often innovations without R&D and, therefore, almost never patented or even published or cited. Hence patent counts and citation counts are a very limited measure to account for knowledge spillovers from users.

When testing the firm size hypothesis, two different effects have to be considered: First of all, a productivity increase with firm size can arise due to large firms adopting new technologies earlier. Additionally, they have more resources at their disposal to hire and retain high quality researchers. A second reason for this relationship may be that R&D is organized differently in large firms. Possibly, scientists in large R&D departments play a smaller role in any single R&D project but are involved in more projects at the same time (Kim et al. 2004). It is assumed that R&D, especially in large pharmaceutical and chemical firms, is organized as proposed by Kim et al.. Inventors are highly specialized and work as members of large inventor teams, on many different projects in parallel. Whereas the first effect represents the Accumulated Advantage Hypothesis, the second effect represents organizational differences leading to biased results. In order to control for these organizational differences, fractional patent counts were used, which turned out to work quite well. Firm size coefficients that are still significant have a value which is a third compared to the results using whole patent counts. These coefficients, finally, are assumed to present productivity differences due to the availability or assignment of resources.

Age was included in the regression to estimate a coefficient for age instead of taking a directly proportional relationship between adjusted patent counts and age (or citation counts and age) for granted. A coefficient of approximately -0.4 implies that the number of adjusted patents/the number of citations rises less than directly proportional with age. This does not mean that productivity steadily increases with age over the whole lifetime of an inventor. Rather Figure 2 revealed an inverse U-shaped relationship. The coefficient of -0.4, therefore, only provides information on the productivity - age relationship before the maximum is reached. Further research should make use of the time structure of the data and look at this relationship more closely.

Further control variables regarding the legal status of the inventors' patents reveal a negative relationship between the share of patents that were opposed by a third party and the productivity index. A possible explanation is that firms active in technical fields, which are characterized by high opposition rates, for instance, the cosmetics industry (Harhoff/Hall 2003), give up business in hard-fought areas, resulting in lower patent counts on the part of the inventors. The share of patents either refused by the EPO or withdrawn by the applicant seems to have a positive effect on productivity. A possible explanation for these results may well be that firms do patent each and every invention that meets the requirements for patentability (novelty, inventive step and commercial applicability). These "lower quality" patent applications are than rejected by the EPO during examination procedure. A withdrawal by the applicant takes place in case the firm is either no longer interested in obtaining a patent for the underlying invention or forestalls rejection by the EPO.

Last but not least, some limitations of this study should be mentioned. First of all, the applied productivity measures are based on patent counts and citation counts. These measures should be a good proxy for productivity in industries characterized with a high patent rate. Griliches (1990) constitutes the fact that "not all inventions are patentable, and not all inventions are patented" as one of the disadvantages of patent data used as an output measure. Further research should include inventions that have been kept secret.

Secondly, the analysis is based on EP data. Since the European patent system was only established at the end of the 1970s, the productivity of inventors who were active in the late 1970s and early 1980s is at risk of being underestimated. In this context, including national patent data from the German Patent and Trademark Office could be very helpful to reduce this bias. Therefore, national patent data will be included in an upcoming version of this paper.

Finally, this paper estimates inventor productivity assuming that inventors kept on inventing during the whole time period under consideration. Controlling for inventors who left R&D for another job, for instance, for administration, is not possible.

The mentioned limitations have to be taken into account when deriving implications from the results. Nevertheless, the analysis provides a deeper insight into the determinants of inventor productivity: overall, both, the allocation of resources and the inventive ability are important determinants of inventor productivity. For R&D management, these results imply that hiring a star-inventor without providing the necessary resources will not suffice to maximize inventive output. On the other hand, providing a maximum of resources will not turn an average inventor into a star-inventor.

References

Afuah, A. (2000). How Much Do Your Co-opetitors' Capabilities Matter in the Face of Technological Change?, *Strategic Management Journal* 21 (Special Issue): 387-404

Allen, T. (1977). Managing the Flow of Technology, Cambridge, MA.

Allison, P.D. / Stewart, J.A. (1974). Productivity differences Among Scientists: Evidence for Accumulative Advantage, *American Sociological Review* 39(4): 596-606.

Arora, A. / David, P.A. / Gambardella, A. (1998). Reputation and Competence in Publicly Funded Science: Estimating the Effects on Research Group Productivity, *Annales D'Économie et de Statistique 49/50: 163-198.*

Bayer, A.E. / Folger, J. (1966). Some Correlates of a Citation Measure of Productivity, Science, *Sociology of Education* 39: 381-389.

Becker, G.S. (1964). Human Capital, New York, NBER.

Blind, K. / Edler, J. / Frietsch, R. / Schmoch, U. (2003). Erfindungen kontra Patente – Schwerpunktstudie zur technologischen Leistungsfähigkeit Deutschlands, Fraunhofer Institut für Systemtechnik und Innovationsforschung, Endbericht, Karlsruhe, December 2003.

Cohen, W. M. / Levinthal, D.A. (1989). Innovation and Learning: The Two Faces of R&D, *The Economic Journal* 99 (397): 569-596.

Cohen, W. M. / Levinthal, D.A. (1990). Abvsorptive Capacity: A New Perspective on Learning and Innovation, *Administrative Science Quarterly* 35: 128-152.

Cole, J.R. / Cole S. (1973). Social Stratification in Science. Chicago University Press.

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

Denison, E.F. (1964). Measuring the Contribution of Education, *The Residual Factor and Economic Growth*, OECD: 13-55.

De Solla Price, D.J. (1965). Networks of Scientific Papers, *Science* 149(3683): 510-515.

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.

Foray, Dominique (2004). The Economics of Knowledge, Cambridge: MIT Press.

Freeman, C. (1991). Networks of Innovators: A Synthesis of Research Issues, *Research Policy* 20: 489-524.

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.

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* 28:1661-1707.

Guilford, J.P. (1968). Intelligence Has Three Facets, *Science* 160(3828): 615-620.

Hall, B.H. (2004). Exploring the Patent Explosion, UC Berkeley und NBER Discussion Paper, Berkeley.

Hall, B.H. / Mairesse, J. / Turner, L. (2005). Identifying Age, Cohort and Period Effects in Scientific Research Productivity: Discussion and Illustration Using Simulated and Actual Data on French Physicist, paper presented at the Keith Pavitt Memorial Conference, SPRU, November 2003, revised June 2005.

Harhoff, D. / Hall, B.H. (2003). Intellectual Property Strategy in the Global Cosmetics Industry, unpublished manuscript, University of Munich.

Harhoff, D. / Wagner, S. (2003). Modeling the Duration of Patent Examination at the European Patent Office, unpublished manuscript, University of Munich.

Idson, T.L. / Oi, W.Y. (1999). Workers are More Productive in Large Firms, *American Economic Review* 89(2): 104-108.

Jaffe, A.B. (1989). Real Effects of Academic Research, in *American Academic Review* 79 (5): 957-970.

Johnson, J.M. / Regets, M. (1998). International Mobility of Scientists and Engineers to the US – Brain Drain or Brain Circulation? *NSF Issue Brief* 98-316, June, 1998.

Jones, B.F. (2005). Age and Great Invention, NBER Working Paper No. 11359, Cambridge, M.A.

Jovanovic, B. (1979). Job-Matching and the Theory of Turnover, *Journal of Political Economy* 87(5): 972-990.

Kim, J. / Lee, S.J. / Marschke, G. (2004). Research Scientist Productivity and Firm Size: Evidence from Panel Data on Inventors, Working Paper.

Kwok, V. / Leland, H. (1982). An Economic Model of the Brain Drain, *The American Economic Review* 72(1): 91-100.

Lehman, H.C. (1966). The Most Creative Years of Engineers and Other Technologists, *Journal of Genetic Psychology* 108: 263-277.

Liu, P.-W. (1986). Human Capital, Job Mobility and Earnings Growth between Jobs: An Empirical Analysis, *Applied Economics* 18: 1135-1147.

Los, B. / Verspagen, B. (2003). Technology Spillovers and Their Impact on Productivity, Working Paper.

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.

Narin, F. (2000). Tech-Line Background Paper, in: J. Tidd (Ed.): *Measuring Strategic Competence*, Technology Management Series, London.

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

Oberg, W. (1960). Age and Achievement – and the Technical Man, *Personnel Psychology* 12: 245-259.

Pelz, D.C. / Andrews, F.M. (1966). Scientists in Organizations: Productive Climates for Research and Development, New York.

Sher, I.H. / Garfield, E. (1966). New Tools for Improving and Evaluating the Effectiveness of Research, in: M.C. Yovits / D.M. Gilford / R.H. Wilcox / E. Staveley / H.D. Lemer, (Eds.): *Research Program Effectiveness*, New York: 135-146.

Shockley, W. (1957). On the Statistics of Individual Variations of Productivity in Research Laboratories, *Proceedings of the Institute of Radio Engineers* 45: 279-290.

Topel, R.H. / Ward, M.P. (1992). Job Mobility and the Careers of Young Men, *Quarterly Journal of Economics* 107: 439-479.

Turner, L. / Mairesse, J. (2003). Individual Productivity Differences in Scientific Research: An Econometric Study of the Publications of French Physicists, Discussion Paper.

Vincent, H.F. / Mirakhor, A. (1972). Relationship between Productivity, Satisfaction, Ability, Age, and Salary in a Military R&D Organization, *IEEE Transactions on Engineering Management* 19(1): 4-52.

Von Hippel, E. (1988). The Sources of Innovation, Oxford.

Widerstedt, B. (1998). Moving or Staying? Job Mobility as a Sorting Process. UMEA University Dissertation.

Zuckerman, H. / Merton, R.K. (1972). Age, Aging, and Age Structure in Science, in: R.K. Merton (Ed.), The Sociology of Science: Theoretical and Empirical Investigation, Chicago, 497-527.

ANNEX 1: MATCH OF ADRESSES

% OF MATCH	LAST NAME	FIRST NAME	FIRST NAME (PART)	STREET	STREET (PART)	STREET_2	CITY	CITY (PART)	CITY_2	APPLI-CANT	FREQUENCY	PERCENT
100%	x	x		x			x				19,487	65.02
100%	x		x	x			x					
100%	x			x			x					
100%	x	x				x			x		6,769	2.26
100%	x		x			x			x			
100%	x					x			x			
99%	x	x		x				x			2,200	7.34
99%	x		x	x				x				
99%	x			x				x				
97%	x	x			x		x				76	0.25
97%	x		x		x		x					
97%	x				x		x					
95%	x	x					x			x	0	0.00
95%	x		x				x			x		
80%	x	x					x				4,185	13.97
80%	x	x							x		0	0.00
INVENTORS QUESTIONNED IN THE PATVAL SURVEY											3,049	10.17
INVENTORS WHO FILLED OUT TWO OR MORE QUESTIONNAIRES											297	0.99
SUM											29,971	100.00

PATVAL DATA

- The PatVal data include information on 3,346 granted EP patents with priority date between 1993 and 1997. Since 297 inventors filled out two questionnaires, the dataset contains 3,049 different inventors who form the basis for the following search procedure.

SEARCH DATABASE

- Use of the EPOLINE patent database of the European Patent Office (EPO). The dataset is a counterpart of the EPOLINE® database as of March 1st, 2003 and covers over 1,260,000 patent files with application dates ranging from June 1st, 1978 to July 25th, 2002. Data on inventor addresses were available until 1999. The search procedure aims at identifying all EP patents of each inventor contained in the PatVal data.
- In order to do this, co-inventors were split. For each inventor duplicate records were created, leading to 2,436,260 inventor files. In order to limit the search effort, last names not included in the PatVal data were removed from the dataset. Due to this procedure, the inventor files could be reduced from 2,436,260 to 231,681 inventor files.

DATABASE QUERY
The query was carried out using MYSQL version 4. The MYSQL-control center was applied as SQL-Interface. All Java classes were constructed with Eclipse.

ADJUSTMENT BEFORE SEARCH

Both datasets, the PatVal data (including 3,049 inventor files) and the EPOLINE data (including 231,681 inventor files), were standardized as follows:

- ß → ss
- ä → ae
- ö → oe
- ü → ue
- removal of “;“ and “,“
- use of small letters (Meier, Hans-Juergen → meier, hans-juergen)
- harmonization of the names’ spelling, of street names or city names
- split of last name, first name and title
- split of zip code and city name
- delete of information concerning post office boxes from address data
- delete “c/o …“ from address data

SEARCH PROCEDURE

- removal of „str., strasse“ from street names
  leopoldstrasse / leopoldstr. / leopold-strasse / leopold-str. → leopold
- removal of house numbers

- removal of corporate form identifiers
  siemens AG → siemens
  microsoft deuschland GmbH → microsoft deutschland
- delete of city add-ons
  frankfurt am main → frankfurt
  st. poelten → poelten
- delete of name add-ons
  stefan jr. → stefan
  stefan I. → stefan
- additional search for parts of a first name, e.g., hans juergen, hans-juergen, hansjuergen, hans, juergen
- additional search for parts of a city name, e.g., bergisch gladbach, bergisch-gladbach, bergischgladbach, bergisch, gladbach)

UNAVOIDABLE MISTAKES RESULTING FROM THE SEARCH PROCEDURE

- brothers jointly owning a firm who assign the firm's address as the inventors' address → last name, street, and city match completely → a search procedure not accounting for the first name results in an additional inventor in the dataset
- married couple with matching last name and address
  → last name and home address match completely → a search procedure not accounting for the first name results in an additional inventor in the dataset
- inventors with common last names employed with large firms (e.g. SIEMENS), assigning the firm's address or the address of a patent agency (e.g. PHILIPS patentverwertung) as inventor's address
  → last name (probably even the first name), street, city, and applicant match completely → in case the first names do not match, the search procedure results in an additional inventor in the dataset. In case the first names do match, information including IPC classes as well as the names of the co-inventors have to be employed to distinguish between PatVal and non-PatVal inventors
- further mistakes are misspellings, e.g., missing hyphens hansjuergen/hans-juergen), ph vs. f. (rudolph/rudolf), v vs. f (detlev/detlef) or inconsistent abbreviations of first names (hans peter/hans p.)
  → due to the additional search for parts of the inventor names, the search procedure detects misspelled inventor names. Unfortunately, the detected patents are not assigned to one but to a number of inventors, resulting in a biased frequency measure:

APPLICATION NUMER	LAST NAME	FIRST NAME	FREQUENCY	ADJUSTED FREQUENCY
EP19890904071	reinartz	hansdieter	1	7
EP19900907047	reinartz	hans-dieter	1	7
EP19900908936	reinartz	hans dieter	1	7
EP19900910736	reinartz	hansdieter	1	7
EP19910912570	reinartz	hansditer	1	7
EP19910903346	reinartz	hans	1	7
EP19910905775	reinartz	dieter	1	7

To minimize the described mistakes, the matching of the inventor files was subsequently checked manually.