Politecnico di Torino

Corso di Laurea in Ingegneria Gestionale

Tesi di Laurea Magistrale

STAR AND PROLIFIC INVENTORS: EMPIRICAL ANALYSIS OF DIFFERENCES BETWEEN PATENTS

Relatore:

Federico CAVIGGIOLI

Candidato

Giovanni SCONOSCIUTO

December 2019

Contents

Acknowledgments

A great thanks to all those who shared all the path with me. Starting from Edoardo and Tommaso, moving to Antonella, Simone, Francesca and Marcello, arriving to my work colleagues. I would like to thank the supervisor of this work, prof. Federico Caviggioli for his continuous presence, his incessant support and infinite availability that allow me to overcome the last months of the path.

A thank to my parents, distant but always present, thanks for your patience, help and support.

Finally, a special thank to Eleonora, without whom it would not have been the same journey. Thanks for making it bright when I saw it dark.

List of Figures

List of Tables

Abstract

Starting from the basic topics of invention, innovation and inventors, the last group will be further analyzed by introducing two categories: “Star Inventors” and “Prolific Inventors” depending on the participation as a finalist or winner at the EPO Award competition. Chapter 2 of the thesis summarizes a description of the current literature regarding these two categories of inventors and the main literary achievements. As the final aim of the work is understanding the differences between these two groups of inventors by comparing the features of the patents filed by each of them, an overview of the main social prizes is presented with particular interest to the EPO Patent Award. According to this point of view, an accurate description will follow of Intellectual Property rights focused on Patent and Patent Application with the aim of providing the reader with the main technical knowledge needed to understand the following analysis. Chapter 3 mainly deals with the method used to gather inventors’ patent data. It describes the use of the Derwent Innovation software, the method of exporting patents’ data using the appropriate query and the main issue we had to face: the homonymy. A description of the disambiguation technique used to solve that problem is given. Chapter 4 gives an overview of inventors’ biographical, educational and career statistics (e.g. accomplishment of a Bachelor degree, a Master degree, a PhD, abroad study experience, international work experience, experience as an entrepreneur, working experience in a company or in a university). These variables will be important because they are control variables of logit regression analysis of chapter 6. Chapter 5 introduces the reader to the main patent variables used to demonstrate the existence of a difference between Star and Prolific Inventors. For each of them the mean and standard deviation have been calculated for both the Star and Prolific inventors (the final considered sample counted 63.971 patents) . Finally a t-test has been completed in order to understand if the found variables can be considered significantly different. These variables will be the independent variables of the multivariate logit regression. Chapter 6 gives the reader the results of the regression analysis. In order to understand if the relationship between the dependent variable and the variables of interest is Non-Significant, Significant or Very Significant, more than one model has been tested. In particular eight models have been implemented in order to avoid the problem of correlation between the independent variables.

Chapter 1

Introduction

Spread by Schumpeter, the concepts of innovation and invention have seen a raised interest. Academics but also universities, research institutions and companies have tried to understand how breakthrough inventions raised and managed to conquer the market. Few works have shifted the interest to the authors of these inventions. This thesis will investigate if there are any distinctions among inventors who are the creators of different great inventions. In particular, the work is based on the idea of the existence of two categories of inventors: the first is the Star Inventors sample, those who have been winners or finalists at the EPO Award competition; the second is the Prolific Inventors sample including those who patented many patents, none of them reaching scientific and commercial success. Starting from these two groups of inventors, the dissertation can be divided into three parts: the first phase of literary studying where we understood how previous papers related to the two aforementioned categories of inventors; the second of data collection through the use of specific software and databases, and the third of data analysis through statistical tools. In particular, during the first phase different fellows concerning the theme of Star and Prolific inventors have been analyzed. None of them reported our inventors' distinction. Furthermore, an interesting trend has been found: during the first year of study of this topic, researchers tried to understand the impact of a particular inventor in a specific field of study, then they shifted their interest to the general study of some important inventors, focusing on their production, their education and collaboration. During the second phase of this thesis, starting from a group of inventors, two sub groups of Star and Prolific inventors have been randomly selected. For each of them his/her patents' portfolio has been created thanks to the use of a specific software containing each inventor's information. The data gathering has not been simple and immediate due to homonymy problem. Some assumptions have been made and, in case of misleading information, the concerned inventor has been excluded from the selected sample. Once the data collection was completed and the different authors' database grouped in a single one, the statistical analysis has been done. The final aim of the work is, first, to understand if the different information one can retrieve from inventors' patents can determine if an inventor can be classified as Star or Prolific. Secondly, to derive impact and quality implication of inventors' patents with respect to his/her output dimensions. In order to do so, a statistical analysis of patents information has been done and then a logit regression test has been used in order to understand the existence of a real correlation between the analyzed variables and the dependent one (being a Star or a Prolific inventor).

1.1 Thesis structure

The thesis is divided into 6 Chapters:

- Chapter 2 introduces the two categories of star and prolific inventors and gives the reader the main technical knowledge regarding intellectual property rights in order to allow him to understand the following analyses;
- Chapter 3 examines the methodology of database creation, the main issues found and the different assumptions made;
- Chapter 4 explains first the biographical and educational differences between

the randomly selected inventors, and then analyzes the differences of patent information between the two samples of inventors;

- Chapter 5 presents eight models of multivariate regression performed on the variables retrieved from inventors' patents;

- Chapter 6 draws the conclusions and the main contribution of this thesis.

Chapter 2

Theoretical Background

This chapter introduces the reader to the main theoretical concepts of this thesis. It describes the state of the art of the current literature regarding the main topics the work will deal with. Starting from the basic topics of invention, innovation and inventors, this last group will be further analyzed by introducing two categories: "Star Inventors" and "Prolific Inventors" depending on the participation as a finalist or winner at the EPO Award. A description will follow of the current literature regarding these two categories of inventors and the main literary achievements.

As the final aim of the work is understanding the differences between these two groups of inventors by comparing the features of the patents filed by each of them, an overview of the main social prizes is presented with particular interest to the EPO Patent Award. According to this point of view, an accurate description will follow of Intellectual Property rights focused on Patent and Patent Application with the aim of providing the reader with the main technical knowledge needed to understand the following analysis.

2.1 Invention, Innovation and introduction to Inventors

This section contains the most general topics that need to be defined due to their undiminished importance in the economic progress and, besides, they can be considered the starting point of an inventor's idea.

In business terms, invention and innovation are one of the ways (internal to a company) that generates a competitive advantage by undercutting the edge of the previous market leaders. This concept is the foundation of Schumpeter's idea of *"Creative Destruction" that marks out "the process of industrial mutation [...] that incessantly revolutionizes the economic structure from within, incessantly destroying the old one, incessantly creating a new one."* [1] Joseph Aloïs Schumpeter (1883-1950) has been considered as the economist who integrated the innovation concept in the economic studies by introducing the idea of Innovation Trilogy that divides the technological change process in three stages:

1. Invention: encompassing the generation of new ideas or the idea to do existing things in a new way;
2. Innovation: including the transformation of the previous ideas into marketable products or processes;
3. Diffusion: the spread of the generated product or process into the market.

In other words Schumpeter agrees that an invention becomes an innovation only if, thanks to the production system, it goes through the market regardless of the success raised. Moreover, the idea of "Creative Destruction" is proved by the concept of "Product Life Cycle" [2] shown in *Figure 2.1.* According to this idea, there will be a new invention during the decline phase that will disrupt the previous one. Recently W. Chan Kim and Renée Mauborgne have outlined a new idea of Innovation [3]. The

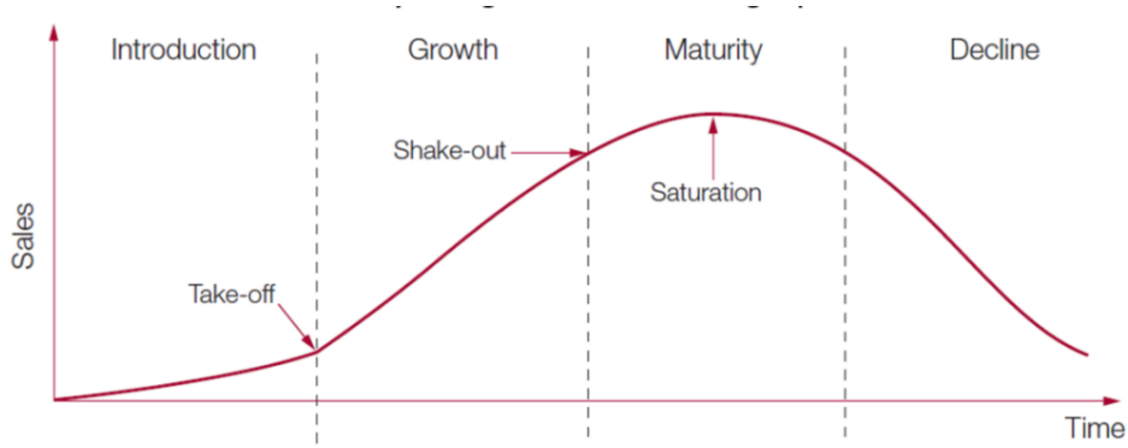

Figure 2.1: The product Life Cycle

"Value Innovation" is the process that brings about the generation of a competitive advantage by improving existing services or products and lowers their costs for both the company and customers. This concept is the basis for a "Blue Ocean Strategy" that focuses on creating new markets instead of fighting against competitors for existing market shares.

Invention and the innovation will not be possible without someone that starts this process thanks to his ideas and knowledge. The Inventor, a single person or a group of people, represents this figure. The inventor usually works as a researcher in a University, firm, Public Research Organization (PRO) or in University spin-offs (USOs).

2.2 Who Star Inventors and Prolific Inventors are

The object of this paragraph is to present the main actor of the innovation process - the Inventor - and the dichotomy between "Star Inventors" and "Prolific Inventors". The current state of the art consists of some studies made by experts that emphasize the importance of some individuals in the process of knowledge creation. The literature that will be presented does not identify a clear distinction of these two categories but introduces the bases of knowledge that allow the dichotomy. In this work, we will refer to "Star Inventors" as the ones that, thanks to the relevance of their inventions, have generated an important benefit to the community from both a scientific and marketability point of view. They are responsible for radical or breakthrough inventions that, according to Ahujia and Lampert [4], are *"those foundational inventions that serve as the basis for many subsequent technical developments"*. On the other hand, we consider as "Prolific Inventors" the authors of multiple unsuccessful inventions that have never left a footprint according to the previous criteria of evaluation.

2.2.1 Literature on Star and Prolific Inventors

As introduced above, according to the current public literature, no work remarks clearly the distinction of the inventors into these two categories. The current section will outline the most important literary references to both Star and Prolific Inventors. During the last 20 years, Biotechnology has been the most productive field that has led to many of the "Star Inventors" definitions. Lynne G. Zucker and Michael R. Darby [5] have given one of the first definition of "Star Inventor" describing them as *"those with more than 40 genetic sequence discoveries or 20 or more articles reporting genetic sequence discoveries by 1990"*. A similar definition has been provided by Niosi and Queenton [6]: *"biotech superstars are those with more than five patents and more than one major publication per year"*. Likewise, for the photovoltaic sector, X. Han and J. Niosi [7] by using the term "Star Scientist" expressed the same concept: *"academic star scientists are university or institutional researchers with at least four photovoltaic industry patents in the USPTO and/or over 100 SCOPUS publications"*. The raising interest – over the time – in Star Inventors has brought about the generation of specific papers, relevant for all scientific disciplines, that are focused on their definition and understanding the importance of their works. First, Tartari defined Star Scientists as *"academics [...] in the top 1% of the distribution of citations in* *their discipline, and the top 25% of the distribution for grants"*. Jan Hohberger [8] introduced the idea of *"standing on the shoulders of giants" that "is often used by economists, sociologists and historians to describe progress in science and technology. At the core of this statement lies the notion of a process in which inventors and scientists develop ideas, based upon the discoveries of other inventors or scientists, and where new ideas add to an existing stock of knowledge. [...] Research in science and innovation has demonstrated that a small group of individuals, often called 'star scientists' or 'star inventors', is associated with generating a disproportionately large amount of scientific and technological ideas. Therefore, 'stars' might be the 'giants' in knowledge development on whose 'shoulders' we stand"*. The paper shows, thanks to a Negative Binomial Regression, that the presence of a Star Inventor or the production of new inventions built on star ideas give them more value. C. Le Bas and R. Bouklia-Hassane [9] in their work, instead, debate on "Prolific Inventors" comparing them with Harhoff's idea of Star Inventors *"inventors who have highly cited patents matching innovations of larger technical and market value"*. In their paper, they argued that star scientists and prolific inventors are highly productive inventors, the former in Science while the latter in Technology. As disclosed in the previous section, in the aforementioned papers, there is not a clear distinction between Star Inventors and Prolific Inventors, as we want to present them in this work; in the last paper, they are even compared. The main features found in literature that define a Star Inventor as such, are linked to the literary productivity along with literary quality that can be assessed from the number of citations, number of patents filed and characteristics of the patents. Summarizing, inventors are considered Stars if they can be plotted in the right tail of a productivity distribution [10] (see *Figure 2.2*).

No specific references to the Prolific Inventors – according to our paper idea – have been found. In view of what has been said about Star and Prolific Inventors,

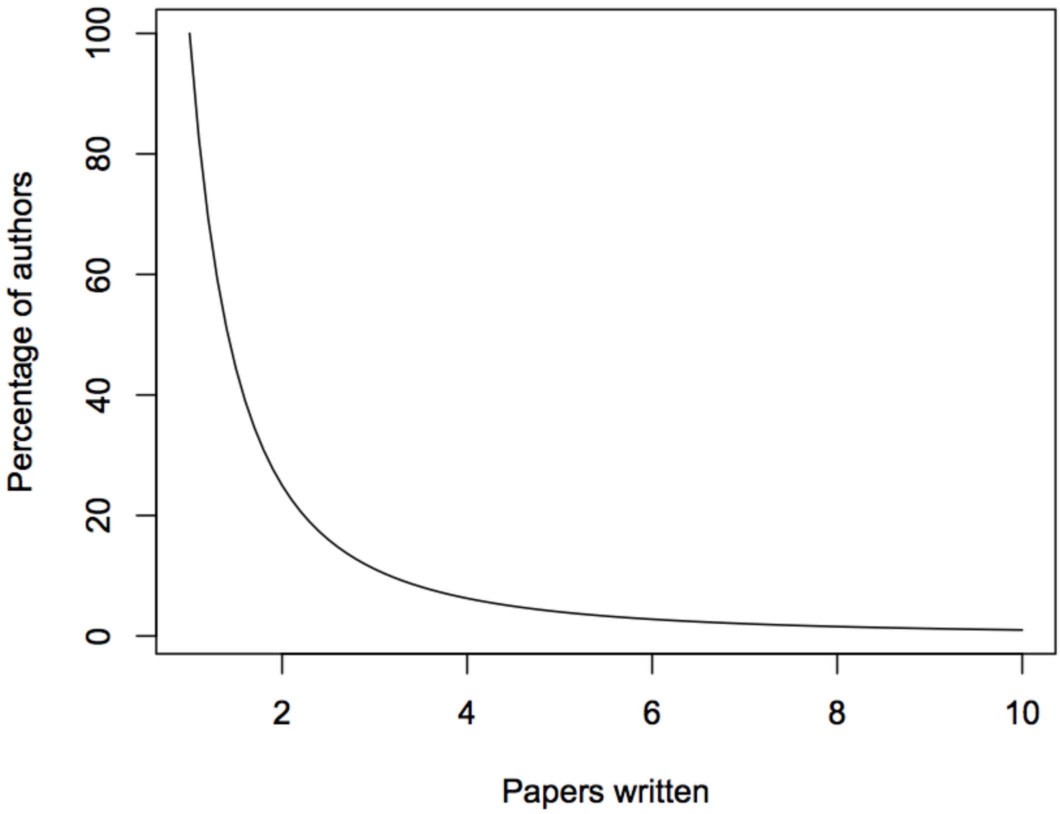

Figure 2.2: The Lotka's function

it is appropriate to introduce our idea of classification of the Inventors into one of these categories, the full explanation of this process will be deeply analyzed in the next chapter. Scientists who have won or participated as finalists in the EPO Patent Award are considered Star Scientists and they are opposed to the control sample of the Prolific Inventors retrieved by the European Patent Office. The aim of our work is to understand if there are some features in their Patent characteristics that contradict our classification of such inventors into one of these categories.

Team working experience

The aim of the current section is to describe briefly the main distinctive traits of Star Inventors' production activity. In particular, the focus is pointed, firstly, on their working collaborations in order to understand if they operate as lone wolves or if they benefit from team working experiences. Secondly, we want to understand if Star Inventors' literary production is focused on few themes or if it ranges on different topics. L.C. Reichensperger in her paper [11] has reported the result of The R&D 100 Award inventor survey: a questionnaire addressed to 4630 inventors that have been awarded by the Research & Development magazine for their breakthrough inventions between 2005 and 2014. The average response rate was 8.96%: 415-awarded inventors have contributed by revealing some information about their invention process. 89% of them have worked in teams made of 0-10 inventors, 7% in 11-20 inventors' team, 2% in 21-30 inventors' team and another 2% collaborated with more than 30 colleagues. The average size of the team of inventors was around 7. Furthermore, the inventors were asked to identify their colleagues' organization. 73% of the collaborations were between colleagues of the same inventor's organization, followed by government research organizations (32%) and universities (19%). The tendency of group working, as reported by Christiane Goetze [12], is also visible by the number of inventors per patent. At the German Patent Office, it has risen from an average of 1.96 in 1995 to 2.32 in 2005. Zhenzhong [13] has obtained similar results concerning other European and American countries. In *Figure 2.3* presented below one can see the average number of inventors per patent in each country where the analysis has been carried out. In general, an increasing trend brings from 1.63 inventors in 1980 to 2.45 in 2005.

Jones [14] has found the same results; he performed a regression analysis relating team size to the patent application year, the time trend found is robust to a number of controls. Some following studies have shifted the focus on the importance

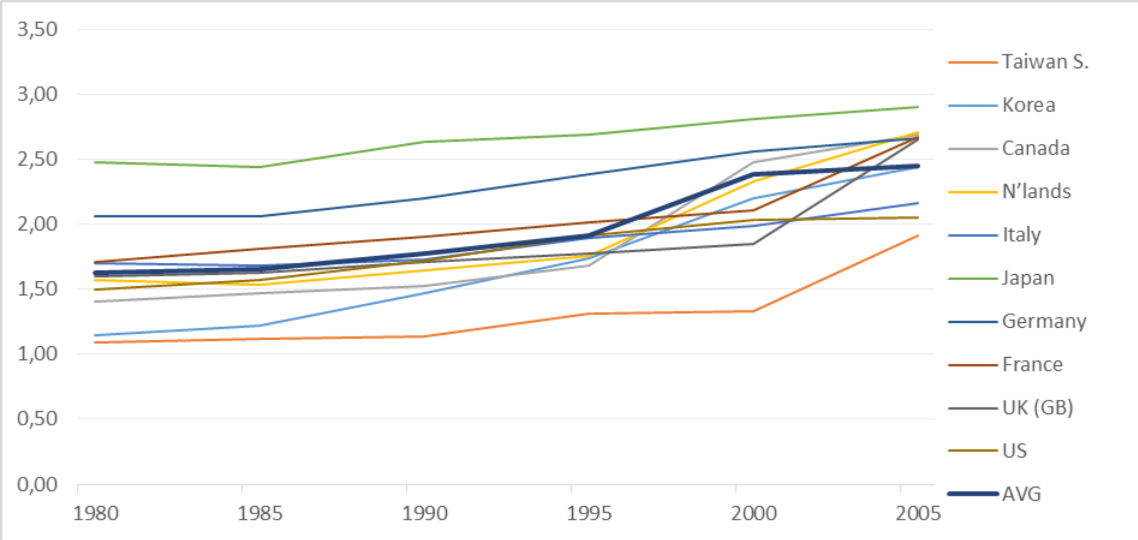

Figure 2.3: Average number of inventor per patent (Elaboration from Patent application and technological collaboration in inventive activities: 1980–2005. Zhenzhong Ma, Yender Lee. ScienceDirect. Technovation 28 (2008) 379-390

Table 2.1: Inventor's Taxonomy

	Average Productivity	High Productivity
Average helpfulness	Non-Star	Lone wolf
High helpfulness	Maven	All-Star

of team working. Lynne G. Zuckera and Michael R. Darbyb [10] found that the collaboration of the star inventor with his team brings about high firm performance in terms of product development and marketability. Oettl [15] was one of the first researchers to study the Star Inventors' influences on their coauthors. The goal of his study was to examine the extent to which different Star types of researcher influenced the performance of their coauthors. He firstly identified a second variable along with productivity in order to qualify stars and non-stars inventors: average helpfulness. Thanks to this new variable, he defined a new taxonomy for Star Inventors as reported in *Table 2.1*. The current table has been used in Chapter 4 in order to understand the distribution of our sample of inventors into these four categories.

Secondly, in order to understand the importance of stars on their coauthors he examined the quality change in coauthors' productivity after a star's death. For deeper specifications about the helpfulness and quality calculations, we refer to the original paper, in the current section we want to show the result of the fixed-effect Poisson estimator. As one could expect, the death of an all-star caused a 16% decrease in coauthors' performance against 14% in case of maven's death while the estimate for lone wolves was statistically insignificant. A similar study [16], focused on collaboration between researchers and Star Inventors, showed that young scholars who have the opportunity to work with a well-respected scientist have greater probabilities to have a successful career. In particular, it has been evaluated that 27% of these people will have a 10 years' career length against only 5% for scientists who have never worked with a star inventor. Particularly interesting is the fact that, working with a star inventor seems to be more prolific after having acquired experience in the sector.

Thanks to these studies, we can draw some conclusions. On the one hand, there are star inventors that use to work and collaborate with some colleagues; while on the other hand, there still are some inventors that work as lone wolves. Although there is this distinction, one can agree that working with a star inventor is a great privilege that can significantly push the careers of those who have this honor. Generally, as shown in the previous graph, there is a rising trend in inventors' group dimensions. This is mainly due to the fact that modern scientific and technological development requires a deeper knowledge of a narrow scientific field. In this perspective collaboration is fundamental firstly because it allows to build an omniscient group and secondly because it promotes cross-fertilization and cross thinking.

Diversification of competences

The aim of this section is to understand if star inventors' production activity is focused on a single scientific field or if, thanks to its inventor's knowledge, it can range on several areas of expertise. According to our researches, few papers face this theme. As one could expect, the ability to vary between different technological fields, positively affects inventors' production. In their paper, C. Le Bas and R. Bouklia-Hassane [9], thanks to a regression analysis demonstrated this last statement. The result of their test is surely correct but, as anticipated in the previous section, the production of new disruptive discoveries is becoming harder. The inventors are now working at the forefront of technology and science spending a significant amount of time to become specialist within a narrow field of expertise. Jones [14] carried out an interesting analysis where he tried to understand the degree of specialization of innovators over time. The specialization variable is obtained from the probability that an inventor switches the invention's field between two consecutive innovations. The analysis was carried out on the inventors reported on the patents issued by the U.S. Patent and Trademark Office (USPTO) between 1975 and 1993. The results of this analysis show a smooth decrease in switching field probability and a corresponding 6% increase in specialization every 10 years.

Summarizing what we have presented so far, we did not find evidence of a clear distinction between star inventors and prolific inventors in literature. Many papers recognize the importance of star inventors in organizations and in the scientific and academic field by highlighting the productivity and quality of their works. We found an increasing trend in group working: inventors use to collaborate with colleagues, we underline that the dimensions of the collaboration group is increasing over time and the figure of lone wolves is probably no more possible. This, in particular, is due to the extensive specialization of knowledge that requires different experts to work together. For the same reason, we showed that inventors are focusing their studies within a single field and this is reflected in their production activity.

Importance of Educational Experience

In this section, we want to report previous studies that tried to understand if there are some educational features that characterize a star inventor as such. The R&D 100 award inventor survey [11] (aforementioned in the previous paragraph) found that 58% of the awarded inventors are between 41 and 60 years old. Generally, the average is around 47 years with a standard deviation of 10. Almost all of the respondents (97%) completed a university education; instead, 65% of them hold a PhD. The most attended universities are Engineering, Hard Sciences and Biomedical and the majority of the inventors came from America (76%) and Europe (15%). These statistics seem particularly interesting if they are compared with those of the PatVal-EU survey [17]. This paper provides information regarding the characteristics of 9017 European inventors who have seen their inventions patented. Almost 77% of the interviewed inventors have a University degree while only 26% of them completed a PhD. The difference between these data is clear. Both the percentages of graduate and doctorate for the awarded inventors are higher than those referred to the general sample of inventors. Even if the difference between these two types of inventors seems to be clear, as reported in the previous sections, there is no clear distinction between them in literature. In the same way, to our research, there is not a unique point of view regarding the importance of the educational path on the output of the invention activity. Ferragatti is the first that tried to find a relationship between educational and career path and being a Star Inventor. As the previous percentages have roughly introduced, the author – thanks to the use of a multivariate regression analysis – demonstrated in her paper [18] a link between these variables. In particular, she has found a robust result between:

- Having a PhD and being a Star Inventor;
- Having a working experience at a University or Research Institution and being a Star Inventor.

In addition, she found a strong correlation between being a Star Inventor and being an entrepreneur or a Start upper. With this thesis, we will now try to understand if there is a correlation between the characteristics of the Inventors' patent and being or not a Star Inventor. Our sample of Inventors has been divided into two categories; we followed the tendency of identifying as Stars those who received a prize. In particular, our Stars will be those who won or participated as finalist in the European Patent Award while the Prolific Inventors are those who filed numerous patents but none of them with economic or social importance. In order to introduce the EPO award we will firstly give an overview of the most important prizes conferred in the scientific field.

2.2.2 The evaluation of inventors' performance through prizes

As anticipated in the previous section, one can use different instruments to recognize the value of inventors' performance. The two most common ones are prizes and patents. In this paper, in order to differentiate between star and prolific inventors we will use a prize: the European Patent Office award. To our knowledge, many researchers try to understand if the inventors' activity is mostly moved by economic or social recognition [18] [17]. The result of these analyses is quite common: social motivation such as contributing to progress in economics, science and technology seems very important; also, the recognition from peers and colleagues inside and outside the organization is a strong incentive to invent; this can raise an inventor in the research community as a Star. Career advances and monetary reward are important but not essential for inventors. The use of prizes to encourage and incentivize researches and developments and to award disruptive technologies is not of primary importance but it has a long tradition. The first application in the modern era was the Longitude Rewards in 1714 by the British government. The price was intended for those who succeeded in tracing the location of ships on the sea by knowing their latitude and determining their longitude. Today, the most important reward is surely the *Nobel Prize* conferred to the authors of outstanding contributions for humanity in chemistry, literature, peace, physics, physiology or medicine, and economic sciences. Other important awards in the scientific sector are the *Abel Price* and *Fields Medal*, the first and second most important awards in mathematics; the Turing Award recognized to the most important contribution in the computer field. Prizes are also common in the art sector, the *Emmy* and *Grammy Award* for music, television, and the *Academy Award (Oscar)* for film industry.

2.2.3 The European Innovation Award

As aforementioned, in this paper we will divide inventors into the two sub-categories of Star and Prolific Inventors depending on the results of the European Patent Award. We chose the EPO award because it is addressed to Patents, which – as we will explain in the following chapter – thanks to their features, are proxy of innovation activity. *"The European Inventor Award pays tribute to the creativity of inventors the world over, who use their technical, scientific and intellectual skills to make a real contribution to technological progress and economic growth and so improve people's daily lives"*. [19] The competition was launched in 2006 by EPO in order to encourage innovation and to give each inventor the recognition he deserves. As reported in the previous paragraph, inventors are seeking social esteem and thus the EPO award is one of the most prestigious prizes that an inventor can receive. The EPO award is presented in five categories:

- **Industry**: it includes awards given to inventors of breakthrough and commercial inventions patented by large European companies (companies with more than 250 employees and annual turnover more than €50 M);
- **SMEs**: similar to the previous one but it awards the inventors behind the small and medium-size enterprises (companies with less than 250 employees

and annual turnover less than €50 M);

- **Research**: this award is presented to inventors who work in universities and research institutions, thanks to the invention he/she has brought technological progress and reputation for his/her institution;

- **Non-EPO countries**: regardless of the size of the firm, this award is given to the inventors that are from outside the EPO's 38 member states. The invention must be available in Europe and it must have a significant market success;

- **Lifetime achievement**: this category awards the long-term contribution of a European inventor whose efforts have contributed to the growth of his/her technological field.

The process of nominations is simple, everyone can nominate an inventor by filing the nomination form, then the EPO experts and an independent jury select the finalists (from two to four for each category) by evaluating the technological originality and economic and social impact of the innovation. Leading experts in the fields of intellectual property, business, science, politics, media and research make up the jury. In parallel, the public can vote their favorite invention: the one that receives the most online votes wins the **Popular Prize**. Given that the European Inventor award is referred to patents as proxy of social innovation, the following section will introduce the Intellectual Property rights with a focus on patents.

2.3 Patent as proxy of innovation activity

The object of this paragraph is to give the reader the main concepts about Intellectual Property rights in order to understand the process that brings from an invention to a Patent. The focus of our analysis is thus the Patent. The reason that brings someone to patent will be presented and an overview of the alternatives to patent in order to protect an invention will be discussed. The final aim of this section is to examine some patents' features considered as synonymous of innovation and patent's value. This means that not all Patents have the same value; some of them bring breakthrough innovation while others are only used for different and less important purposes (e.g. creating a defensive patent portfolio).

We want now to introduce the Intellectual Property System and our object of interest: the Patent. A focus on the patents' characteristics will be provided because it will be essential in order to understand the econometric analysis of Chapter 4. Intellectual property (IP) is a macro-category that includes all intangible creation of human intellect. The intellectual property rights have been introduced at the end of the 19th century to encourage the creation of intellectual goods and to fix the problem of appropriation of their generated value. Without the IP System, a competitor can benefit from the innovator's R&D investment. In particular, after having invested in developing the new product, establishing the supply chain for production, running marketing campaigns and finding distributors, the competitor can simply benefit from all these efforts being able to offer the same product at a cheaper price. The IP System has been created in order to protect the inventor giving him the rights to exclude the others from the product. A brief description of the most important IP rights follows:

- **Patents**: granted for technical inventions;
- **Utility models**: similar to patents, they offer protection for a shorter period of time;
- **Copyrights**: they automatically protect any type of original and creative expression, including literature, art, drama, music, photographs, recordings and broadcasts without being registered;
- **Trademarks**: names, logos and colours applied to the owner's products or

services, which distinguish them from products and services provided by competitors;

- **Registered designs**: they protect the external appearance of a product;
- **Trade secrets**: they cover information not known to the public.

This paper will focus on Patent as a unit of innovation activity; therefore, a deep analysis on Patents follows. *"A patent confers perfect appropriability (monopoly of the invention) for a limited time in return for a public disclosure that ensures widespread diffusion of benefits when the patent expires"* [20]. For its features, the patent is considered a contract between the applicant and the society that protects both their interest. The former can – personally and uniquely – benefit from his invention while the latter encourages innovation and stimulates continuous improvement. In return for this protection the applicant has to reveal the details of his invention to the public, so others can try to improve it. Not all inventions can be patented, according to the European Patent Convention (EPC) *"European patents shall be granted for any inventions, in all fields of technology, provided that they are new, involve an inventive step and are susceptible of industrial application".* [21] Thus, the Invention has to meet three important criteria: innovation, non-obviousness (inventive step) and industrial application. An invention can be considered "new" if there are not previous public disclosures on it. The "Inventive step" is not easy to assess. In order to understand if an invention includes an inventive step the European Patent Office compares it with what would have been obvious to an imaginary person skilled in the same area according to the art at the time of filing. The last requisite expects that the invention must be "susceptible of industrial application". Therefore, patents can protect inventions that solve technical problems such as chemical/pharmaceuticals substances, industrial processes or methods, products/devices. In order to better understand what is patentable, it is important to consult also *Article 52, 53* of the *European Patent Convention (EPC).*

As come to light in the previous paragraph, not all the inventions are patented. In this section, we will give an overview of the alternative techniques to patent trying to understand when each of them can be used. An alternative to patent is the *Utility Model*. It is an appropriative scheme very similar to the patent but it is not available in every country. It grants the holder the exclusive right to prevent third parties from making, using and offering for sale the invention in the country where the application has been filed for a period going from 3 to 10 years, usually its publication is faster than the 18 months required by the Patent. The main advantage for filing a Utility model instead of a Patent are the speed of registration and the financial benefit. A different method of protection is the *Defensive Publications*. It is a kind of report that describes deeply the technical subject matter in order to place this disclosure in the public domain for free to prevent others (e.g. competitors) from obtaining patent on the present matter. The inventor in this way can continue to use his invention but cannot prevent competitors from doing the same. The last alternative is the *Industrial Secrecy*. This method can be used if the process of reverse engineering is not possible. Usually the secrecy is used when - through the disclosure of the invention – it is possible to invent around, by keeping the invention secret this will no more be possible.

2.3.1 Patent application

In this section, we want to focus on the information that one can retrieve from a patent. We will start by explaining the process of patent application by describing the most important information that the patent applicant must provide in order to file the patent. This information will be fundamental because the following analysis will be based on it. The rights derived from a patent take place after the Patent grant. These rights are active in the jurisdiction where the Patent Application has been filed and then granted. Depending on the importance of the invention and the money availability, the applicant can decide to patent his invention in different countries. The first office where the Application is filed is named Office of First Filing (OFF) while the further are called Office of Second Filing (OSF). Publications of OSFs are equivalent to translations of the first publication of the OFF and it must be noticed that – according to Article 4 of Paris Convention [22] – one can add the right of disclosure in different jurisdictions from the OFF claiming the rights of earlier filings. The same result can be obtained if the applicant applies for a special Patent Application; the Patent Cooperation Treaty (PCT) has effect in 148 jurisdictions and is governed by the World Intellectual Property Organization (WIPO). Usually a Patent Application is published 18 months after the filing date, the PCT can require more time because it has to be reviewed in different jurisdictions. The office of filing has then to evaluate the Application, as noticed above the Invention has to meet the three important criteria of innovation, non-obviousness (inventive step) and industrial application in order to be granted. If these requirements are not respected, the Application will be abandoned or withdrawn and any subject disclosed within will belong to the state of the art and so it can be used by others.

2.3.2 Importance to patent – patent value

As explained in the previous section, the patent activity seems fundamental in order to prevent competitors from getting a free ride on the back of the inventor's creativity. In spite of this, not all the inventions are patented. As highlighted by Fontana in his paper [23], within the Economics of Technical change and Innovation Studies (ETIS) literature, there are three main reasons accounting for the decision of not patenting:

1. The innovation is not patentable maybe because of lack of industrial application;

2. The inventive step is not high enough to be deemed worthy of patent protection;

3. The inventor decided to protect his invention with other alternative strategies.

Furthermore, in his paper Fontana, starting from a dataset of inventions that received the *R&D 100 Award* from the *Research & Development magazine*, tried to estimate patent propensity across different dimensions such as industry, geographical area, organization and type of invention. The share of patented innovations varies considerably across sectors and geographical areas: the highest propensity is in the chemical and pharmaceutical sector and Asia is the state with the highest percentage level. In addition, L.C. Reichensperger in her paper [11] tried to find out the main features of patent propensity. She demonstrated that firms used to patent more than PROs and Universities and that large firms, thanks to their patent portfolio that allows them to spread the patents costs, have a high patent propensity than small firms. The task of assessing the value of patent rights from the patent counts is a particularly difficult one, since the distribution of these values is highly skew. One of the attempts used in literature, to solve this problem is to use a set of variables that can be used to approximate patent value. The use of these variables is particularly common because they can be easily retrieved from the patent application. There follows a description of the most important patent characteristics used in order to determine the intrinsic value of the patent:

- **Forward citations**: the forward citations are those citations that the patent receives from works that are more recent, thus, a high number of forward citations means that the innovation has contributed to the future works. The use of forward citations to capture the value of the invention in literature is quite common. The first that started to investigate the individual value of patent through forward citation were Hall and Trajtenberg [24]. They found that the value of a firm is directly linked to its patent citation and that an average increase of one citation per firm's patent increases by 3% the market value of a firm. Consistently, Bessen [25] demonstrated that an additional citation per patent increases the net value of the patent by 4% to 7%. Further studies [26] [27] [8] demonstrated the same positive relationship between forward citations and patent value. Hohberger indicated that the forward citations are highly skewed and the most cited patents are often used to define high-value inventions;

- **Backward citations**: the backward citations must be filled in the patent application and are those previous works to which the patent refers. There is not a unique idea about the role of backward citations and the importance of their uses. Some authors suppose that an application with a broad scope is associated to a high number of citations that brings about a high patent value. On the other hand, since the number of citations is associated to monetary implications, one can be incentivized to omit some references. In the paper [26], the backward citations are positively correlated with the patent's value.

- **Patent families**: a patent family is a collection of patent applications covering the same or similar technical content. The applications in a family are related to each other through priority claims [28]. The first that tried to express patent value from the family size was Putnam. The same study has been proposed by Harhoff, Scherer, Vopel in their paper R26, they found that patent family has a positive and highly significant explanatory power of the patent value.

- **IPC**: it provides for a hierarchical system of language independent symbols for the classification of patents and utility models according to the different areas of technology to which they pertain [29]. In their paper, [30] Harhoff, Scherer, Vopel found that the number of four-digit IPC classifications does not have any explanatory power.

Chapter 3

Methodology

The distinct trait of this work is the database itself and for this reason in this chapter, the process that brings to the creation of the database - used for the following analysis - will be described. The database construction started from a previous work made by Sara Ferragatti [18]. The author's aim was to find some distinct traits, during an inventor's career and education, which can be used in order to distinguish a star inventor from a prolific inventor. The paragraph will discuss the information available in the starting database in order to understand why it has been useful for the creation of the new database. Then the process of building the new database will be explained with particular focus on the sources employed to gather the information and the main issues come upon that influenced the final structure of the database. Finally, an introduction to the next chapter will be provided, in particular the information contained in the database will be the source for a statistical and econometric analysis.

3.1 The starting database

As disclosed in the previous paragraph, this work started from a set of information collected by Sara Ferragatti in her master thesis. In particular, she built a database containing the main career and education information for two sets of inventors. She analyzed 140 Star Inventors (those who have been selected to participate in the EPO Award from 2010 to 2014) and 114 Prolific Inventors (according to an EPO study, those who made a significant number of patented inventions of secondary commercial and economical importance). For each inventor, the author selected a set of information displayed in the *Table 3.1*.

These sets of information have been fundamental in order to build each inventor's patent portfolio.

3.2 The database construction

In this section, one can understand the process that leads to the construction of the new database, with particular focus on the main complications found. Starting from the 254 main Inventor group analyzed by Ferragatti, we extrapolated a random sample of 123 Inventors equally distributed between the Star and Prolific Inventors. Our final Star Inventors group is made of 63 inventors against the 60 Prolific Inventors.

3.2.1 Main sources employed

Starting from this group of 123 Inventors, the aim of the work was to build for each of them his/her patent portfolio. In order to do this, we used the Derwent Innovation software. Derwent Innovation is a Clarivate Analytics company software. It is a market-leading patent research and analytics platform that gives access to trusted patents and scientific literature. After the login in the site, the patent research tool asked to complete some information about the inventor whose patent we want to find. The software offers two different kinds of patent research mechanisms to interrogate the Derwent Innovation DB: the fielded one that allows making query by using a menu of fields; the expert research engine which allows creating command-line style

Table 3.1: The Inventors' career and education set of information

Name (First Last)
ID
EPO
URL webpage (source1: finalists)
URL webpage (source2: all candidates)
Year
Award category
Rank
Team size
Country
Company / Research Centre
Patent Number
Technical field / Sector
The invention in a nutshell
Societal benefit
Economic benefit
Gender
Date of birth
Year of birth (assumed)
Nationality
Bachelor (BA)
BA_year
BA_field
BA_Macrocategories
BA_University
BA_University_Nation
Master (MA)
MA_year

MA_field
MA_Macrocategories
MA_University
MA_University_Nation
Specialistic Degree
Year of speciaistic degree
Specialistic Degree University
Specialistic Degree field
PhD
PhD_year
PhD_field
PhD_Macrocategories
PhD_university
MBA
MBA_year
MBA_univ
Second Degree
Other Degrees
International work experience
International study experience
First experience as Entrepreneur / startupper / founder of companies
Number of years working AT PRIZE DATE
Industry experience (main sector of work)
Worked in a company
Worked in UNIVERSITY/Research centre
Last Job name
Last Company
Link_info

queries (see *Figure 3.1*). We started our research by using the fielded search engine.

Figure 3.1: The Derwent Innovation search Engine

For each inventor we insert his name and surname divided by the word 'NEAR'. The basilar query used to extrapolate the inventor's patent portfolio was '*Name* NEAR *Surname*'. Thanks to this query, the engine found all the patents where the specified inventor worked. The use of the word 'NEAR' allows the engine to do two researches contemporaneously: *Name Surname* and *Surname Name* to avoid the problem of the order used to upload the name and surname of the selected inventor for the specific patent. In some cases, the inventor has a second name: the mechanism used for this kind of inventors is the same described above: *Name* NEAR *Second-name* NEAR *Surname*.

3.2.2 The homonymy issue

As one can imagine, the main problem found in order to download each inventor's patent portfolio was the homonymy one. The aim of this section is to describe how it has been solved in order to understand the main assumptions used during the research phase. We firstly tried to disambiguate the homonymy issue by the use of the Assignee field in the patent application. By comparing the field of Companies where the inventor worked (retrieved from his/her LinkedIn profile if it was known) we did a first selection of the possible patent applications that could be truly referring

Table 3.2: Ratio between the analyzed inventors and the effectively found ones

	Star Inventors	Prolific Inventors	All
Selected Inventors	63	60	123
Identified and analyzed inventors	60	58	118
Identified and analyzed inventors(percentages)	95%	97%	96%

to the inventor object of our analysis. In most cases, this operation was sufficient to exclude all the patents that were referred to a homonymous inventor. If this was not sufficient, we made a strong assumption and we decided to select only the firms that we were sure the inventor worked in (retrieved from the LinkedIn profile or personal Web Pages). A second assumption has been made regarding the inventor's activity period. We assumed that a person could be filed as inventor in a patent application during or after his/her last year of University. In this way, we avoid the error of wrong attribution of a patent to a younger homonymous. We consider the use of these assumptions as reasonable. There are two main criticisms to our assumptions: the overestimating inventors' patent portfolio and the loss of information due to the not updated inventors' LinkedIn or other personal profiles. Regarding the second problem, we assume that it is a common personal interest of each inventor to keep his profile updated in order to gain visibility. In this way, the risk to lose information is minimum and can be accepted. Despite the aforementioned assumptions, there were some inventors whose identity was misleading that we decided to exclude from our analyses. Starting from the 254 main Inventor group analyzed by Ferragati, 5 of the selected 123 inventors have not been recognized certainly and for these reasons we decided to exclude them in order not to negatively influence our sample. *Table 3.2* summarizes the distribution of the excluded inventors between the two inventors' sample groups.

3.2.3 The DB in detail – The template of the extracted information

The aim of our work is to understand if there are some features in inventors' patent characteristics that contradict our classification of such inventors into one of the two categories of Star and Prolific inventors. The focus of this theses is therefore to analyze the patents' characteristics of each inventor. First of all, we had to decide which information we wanted to export – thanks to the use of the software Derwent Innovation – from a patent application. The *Table 3.3* summarizes the selected fields.

Table 3.3: Patent's exported information

ID
Publication Number
Inventor Count
Assignee Count
Claims Count
Count of Citing Patents
Count of Cited Refs - Non-patent
Count of Cited Refs - Patent
Publication Country Code
Inventor - w/address
Inventor - Original
Inventor
Earliest Priority Year
Application Year
Assignee/Applicant
Assignee - Standardized
INPADOC Family ID
INPADOC Family Members
IPC Current Full (4 Characters)
IPC Class
Number of family members
Number of IPC classes

In order to understand the following analysis on the patents collected, one can now find the description of the main variables analyzed in Chapter 4. As we introduced in Chapter 2, the invention activity is no more a solitary activity but it requires a high level of knowledge that can be achieved through the combination of different profiles. Therefore, the variable *Inventor count* wants to measure the number of co-inventors that collaborated to the origin of that patent. The inventor is not always the owner of the rights derived from a patent. Especially if the inventors are working in a R&D unity of a company, they will not be the ownerships of the inventions; it is the patent assignee that has this power. The variable *Assignee count* analyzes the number of different societies that have collaborated in order to conceive the invention. A typical example of co-assignee can be the collaboration of a specific Company with a University, they both participated for the patent conception (through human and economical resources) and for this reason, they are co-assignee. The name of the assignee is resumed in the *Assignee/Applicant* and *Assignee – Standardized* variables. The first reports the full name of the assignee while the second one standardizes it. The following three variables are very important and they must be filed in the patent application when one wants to ask for a patent. They are:

- *Count of Citing Patents*;
- *Count of Cited Refs – Non Patent*;
- *Count of Cited Refs – Patent*;

One can find these variables with the commonest name of Forward Citations and Backward Citations, we refer to Chapter 2 for their descriptions, the novelty here is the presence of two different kinds of Backward Citations, the first refers to the citations to literary works while the second assesses to patents. The other presented variables resume the list of the inventors (*Inventor-Original* and *Inventor*) and their address (*Inventor - w/address*). This last variable could have been very useful in order to deal with the problem of homonymy. Unfortunately, it has been used in a wrong way, even if an inventor comes from nation X but he/she has worked in nation Y to produce the patent, the variable reported the inventor as he/she comes from nation Y. The remaining variables are linked to the family members (*INPADOC Family ID* for the identification code of the family, *INPADOC Family Members* for the list of the publication numbers of the members of the family and *Number of family members* in order to understand the family size) and to the IPC invention class (*IPC Current Full (4 Characters)* and *IPC Class*).

Chapter 4

Statistical Analysis

In this chapter, the results of the statistical and regression analyses are reported. Two different statistical analyses have been completed; the first reports the main features regarding the inventors' sample biographical, educational and career information, while the second one is mostly referred to the inventors' patent portfolios. The focus of this thesis is the second analysis referred to the inventors' patent portfolio, the first one will be briefly presented because it is useful for the following tests. The statistical analysis has been carried out using *Microsoft Excel* in order to understand the differences between the two samples of inventors. In order to distinguish the two groups of inventors, we will use two abbreviations: STAR referring to the finalists/winners of the EPO Award and PROLIFIC for the prolific inventors' sample.

In particular, as reported in the previous chapter, 5% of the STAR inventors (3 people out of 63) and 3% of the PROLIFIC inventors (2 people out of 60) were found but excluded from the analysis due to the high probability of mistake caused by the homonymy problem. For the remaining inventors no other problems have been found and their career and patent data have been collected. Thus the final analyzed sample is composed of 118 Inventors, 60 belonging to the STAR category and 58 to the PROLIFIC sample. The *Table 4.1* and *Figure 4.1* summarize the aformentioned statistics.

Table 4.1: Number of inventors divided in the two analyzed samples

	Star Inventors	Prolific Inventors	All
Number of Inventors	60	58	118

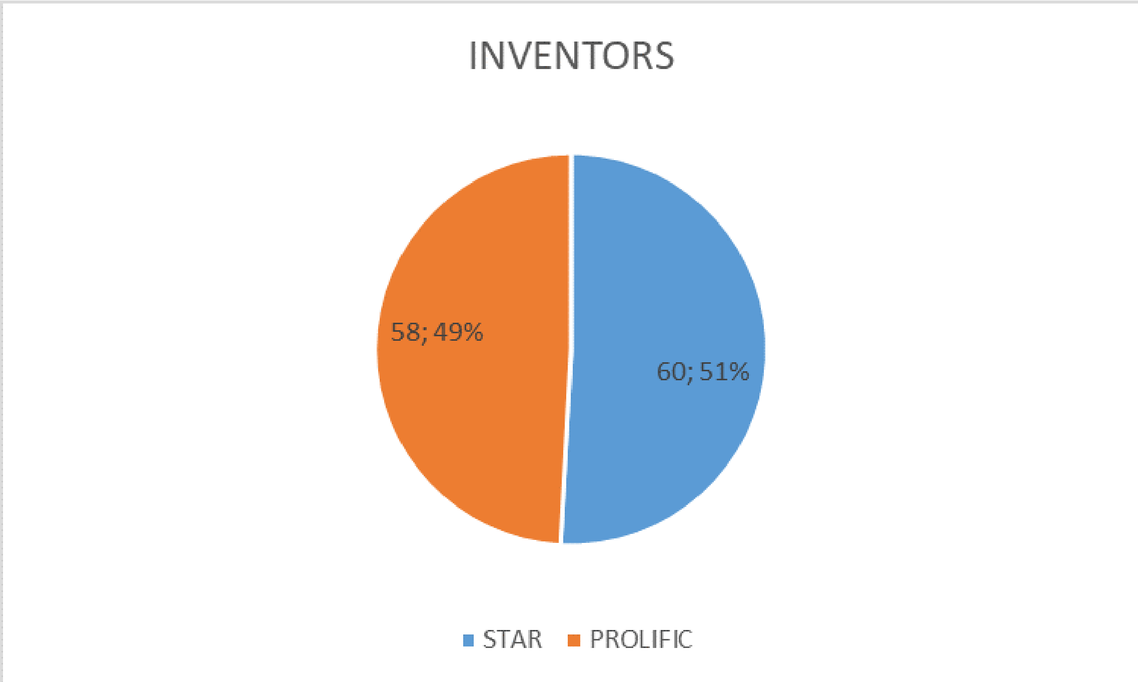

Figure 4.1: Share of analyzed inventors per sample

4.1 Biographical, educational and career statistics

In this section some biographical statistics regarding the gender, nationality and age, some educational statistics (the accomplishment of a Bachelor's degree, a Master degree, a PhD and the possibility to study abroad) and finally some statistics about the career of each inventor (international work experience, experience as an entrepreneur, working experience in a company or in a university) have been analyzed.

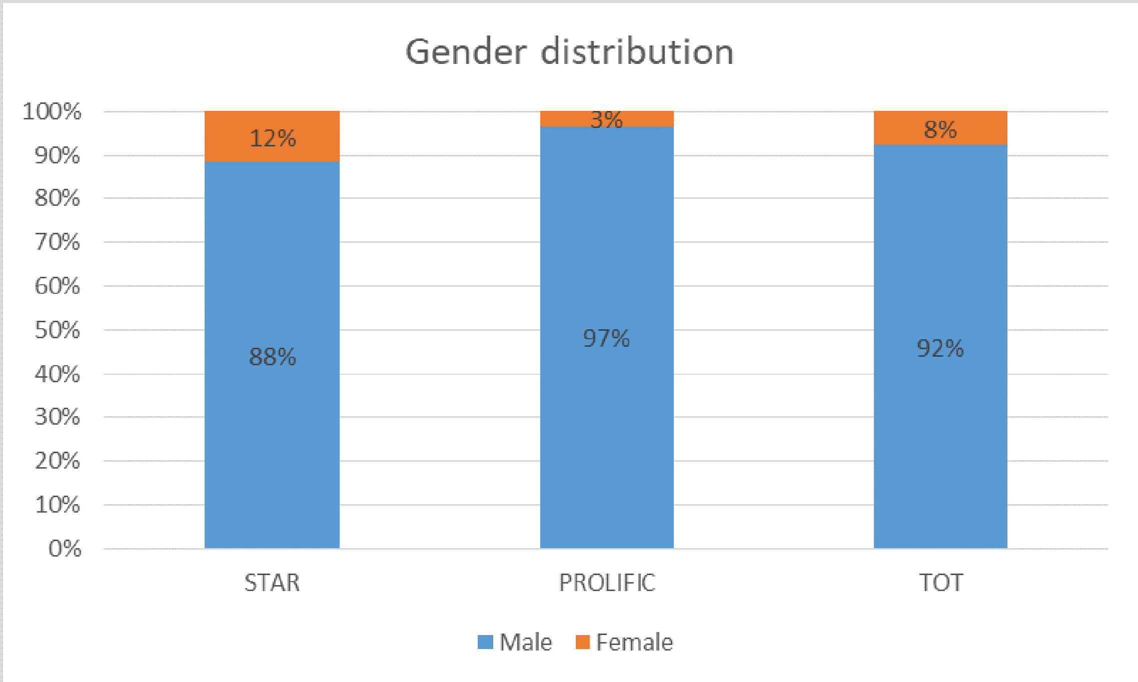

Figure 4.2: Gender distribution of the two samples of inventors

4.1.1 Biographical statistics

Figure 4.2 shows the gender distribution in the two samples of inventors and aggregated. As one can see, only 8% of the overall considered inventors are female and STARS have 4 times female presence than PROLIFIC. We argue that this distribution reflects a known issue on the limited presence women in scientific fields. For example, the European Commission tried to cope with this problem and established rules and incentives to increase the presence of women in the technological and scientific career (see objective "Parity Opportunity" in the program Horizon 2020). Data [31] shows that women in R&D activities are around 45% (30% when considering over 55 years old) and 22% in apical positions. *Figure 4.3* shows the women inventors' share for the top 10 nations from 1998 to 2017. Russia has the highest women percentage in patent applications with 18% while Germany has only 7% of women inventors. France is the first European country to start moving towards 'equal opportunities' male-female in patent field [32].

The following figures (*Figure 4.4*, *Figure 4.5* and *Figure 4.6*) report prolific, star

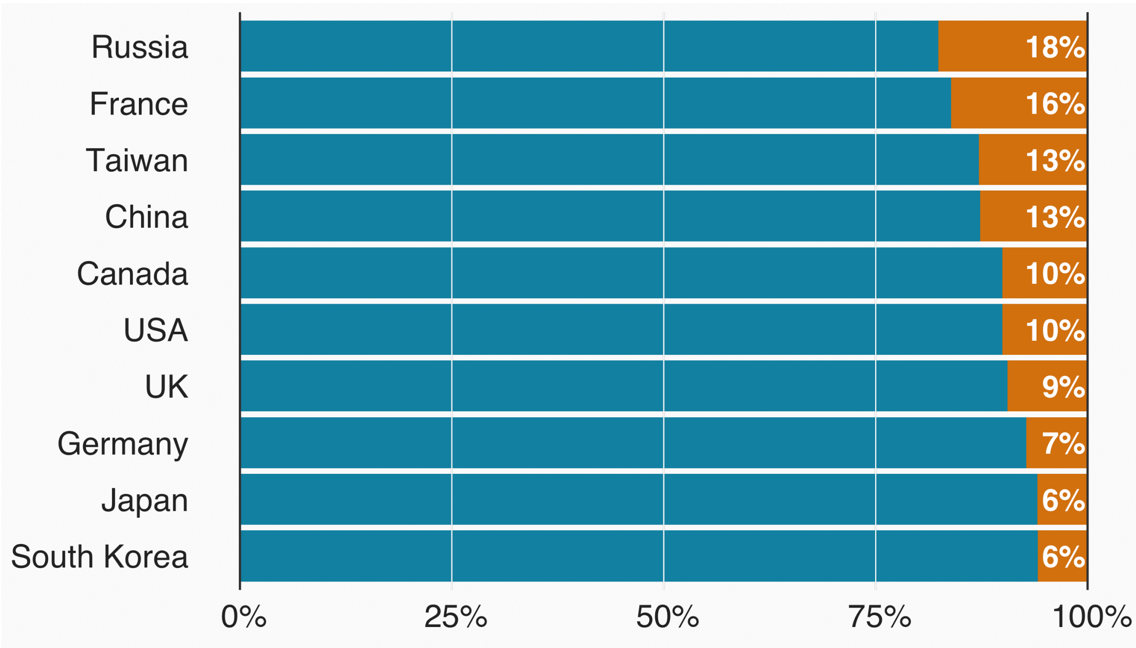

Figure 4.3: Men and Women proportion in patent application, 1998-2017 (Source: IPO analysis of PATSTAT data)

and the aggregation of the overall inventors divided according to their Nationality.

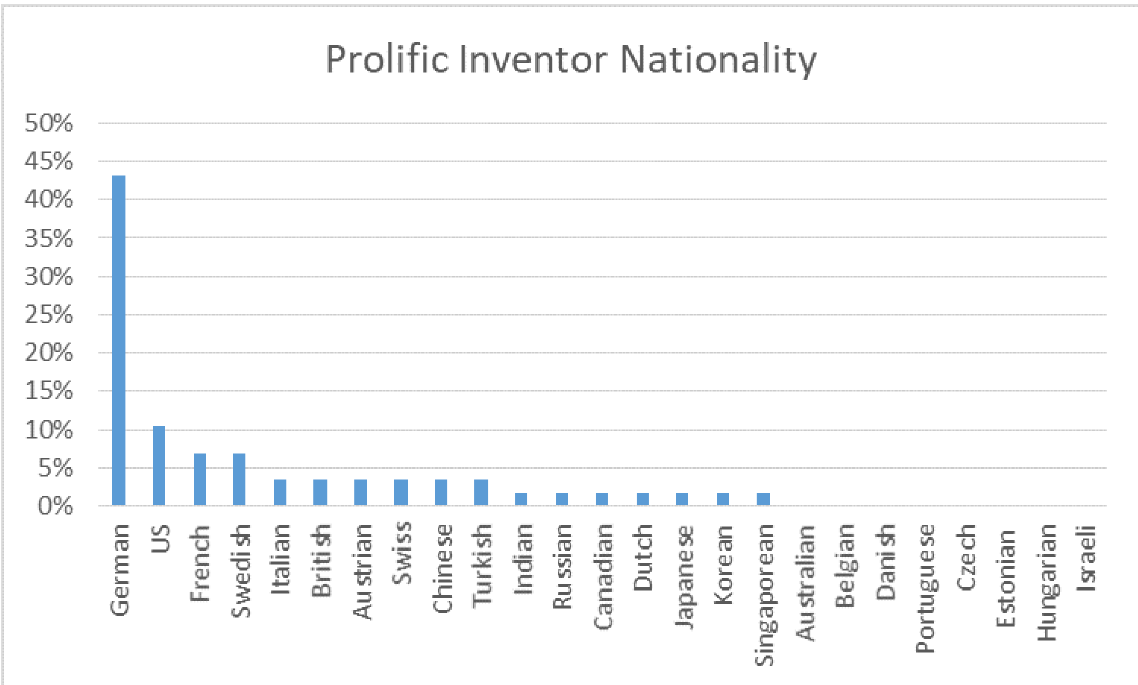

Figure 4.4: Prolific Inventor Nationality

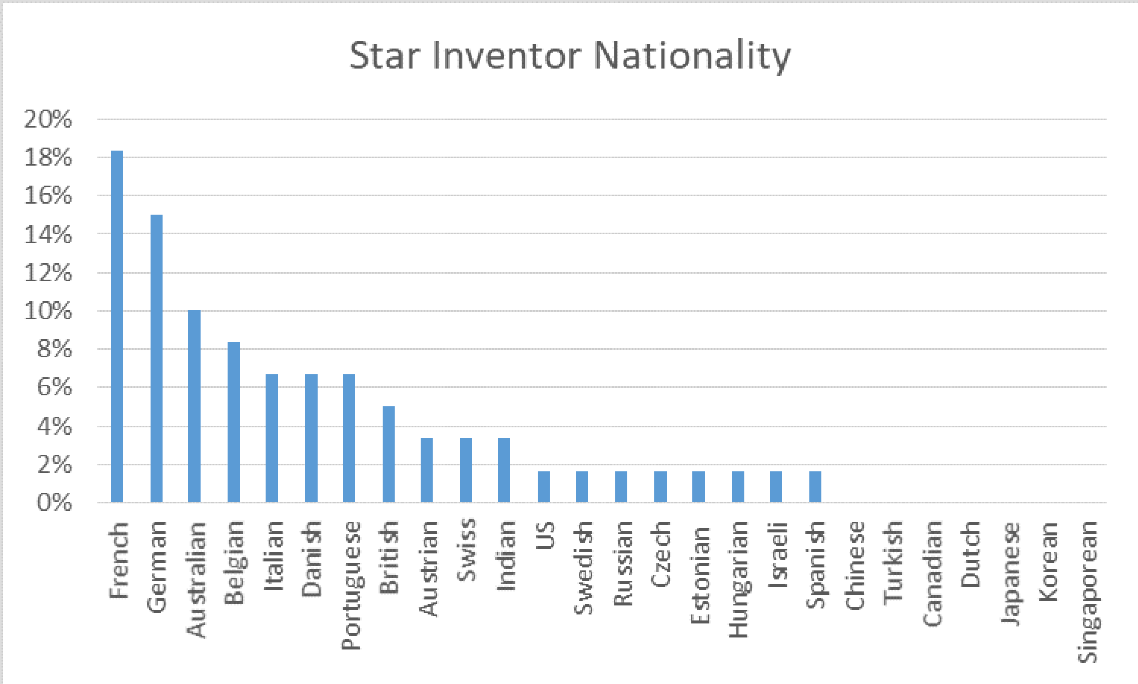

Figure 4.5: Star Inventor Nationality

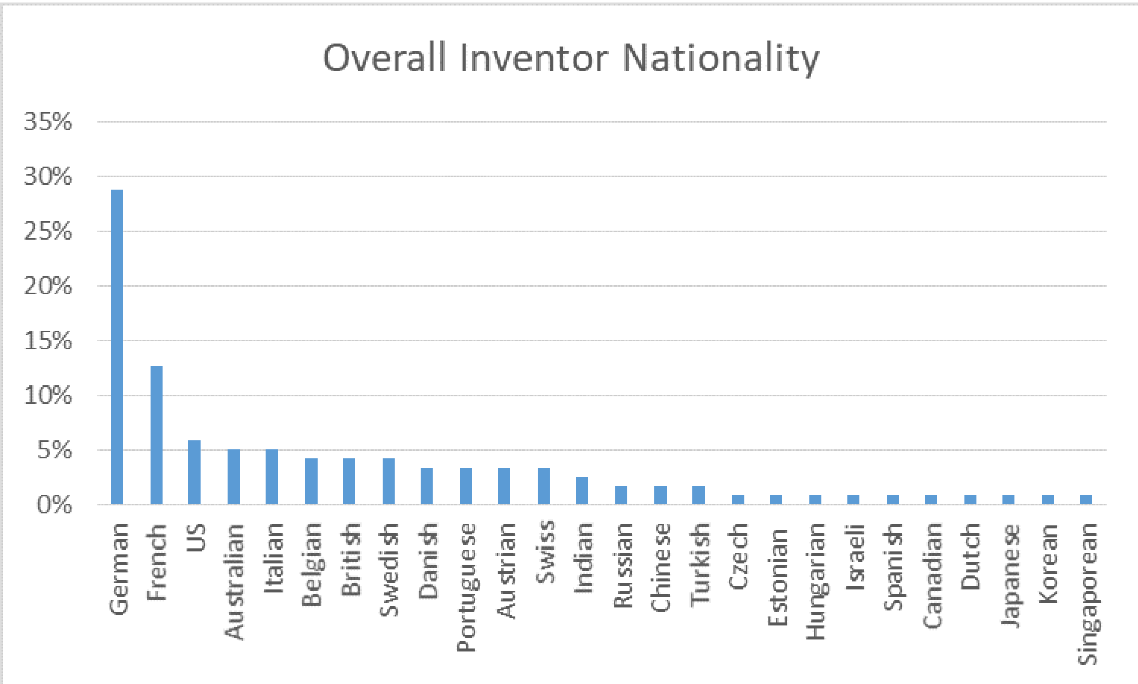

Figure 4.6: Overall Inventor Nationality

The first impressive result is the overwhelming majority of Germans in the Prolific sample, 43% of total that drops to 15% on Star segment. This extraordinary result is mainly coming out from the high investments in Research and Development (Germany allocates 3% of its GDP to R&D investments, almost € 100 billion that exceeds the amount of all other European nations) vs 2,15% of euro zone and 2,25% of France and 1,38% of Italy [33]. Nevertheless, this does not justify the share difference between German Prolific and Stars Inventors. This difference, I believe, is coming out more from a protective culture standpoint and historical capability of Germans to appropriate the returns from Industrial Research and Developments. France is the second absolute country per share of prolific inventors and it is in the first position on Star segment. France has been, historically, one of the first European countries to become aware that research and technological innovation would play a key role in affirming the competitiveness of a country in an increasingly global world. For this reason, it has equipped itself by offering one of the tax incentive systems for Research and Development based on the most attractive tax credit mechanism in Europe. The French system of incentives for Research and Development ("Cir") was introduced in 1983, but has begun to give evident results since it became "volume based" in 2008, that is, proportional to the volume of spending on R&D, instead of being "incremental based" (i.e. calculated only on the increases compared to previous tax periods). Currently the "Cir" is equal to 30% on the first 100 million euros of R&D expenditure, in addition to 5% for expenses exceeding 100 million and without maximum limits. The accrued credit can be carried forward for three financial years and, if not used, it can become repayable: in this case it turns, in fact, into a real cash contribution with consequent possibility, under certain circumstances according to international accounting standards, to account for it "above the line", or between revenues and not to reduce taxes. Spending few words for Italy, our country ranked at the fifth place in the two categories and, this result exceeds the very low investments and poor tax benefits compared to EU. The private sector, businesses and non-profit institutions, is in fact the main source of financing for R&D (Research and

Development) spending, contributing to cover more than 50% of the total amount (€ 21,6 Billion). It is, in practice, € 13,1 billion that serve to activate one of the most important levers to enable our country to compete on the globalized market. Businesses know this well and are even more aware of it; EUROSTAT reported that the Italian budget for R&D has grown from 1,76% of GDP of 10 years ago to 2% of today [34].

The last biographical statistic taken into account is the Age of the inventors. *Figure 4.7* shows the inventors' age distribution for the two samples analyzed.

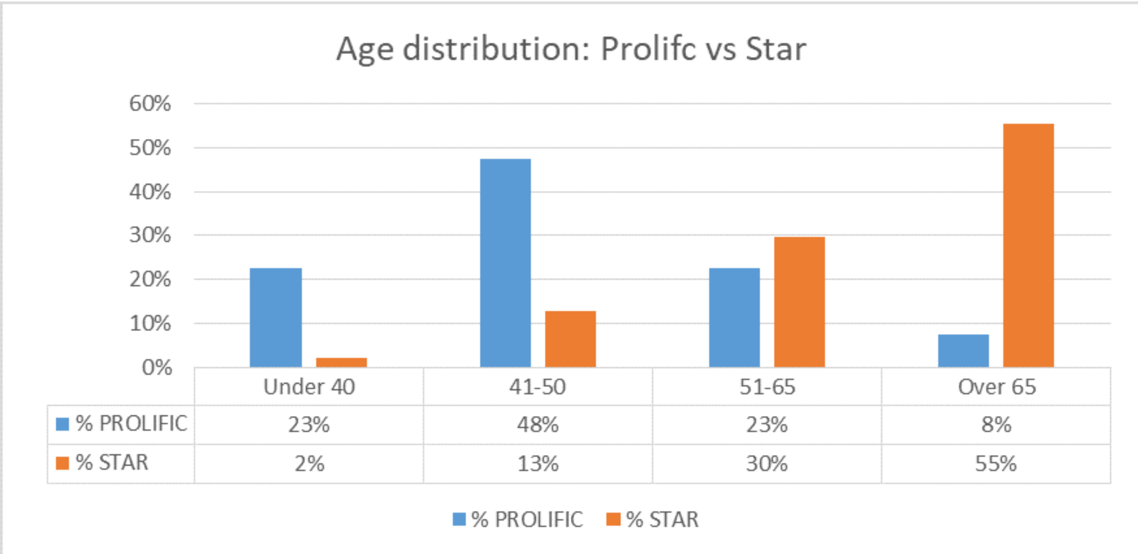

Figure 4.7: Inventors' age distribution

Before describing the two distributions, a focus is fundamental. For the 118 inventors taken into account the age data was available only for 87 of them, for 26% of them it has been impossible to find that information. *Figure 4.7* shows a very interesting result: the two samples have a completely different age distribution, prolific have a Gaussian distribution, centered on 41-50 while Star have an exponential distribution with few inventors under 40 (2% of the sample) and most of them is over 65 years old (55%). No matter of discussion that EPO is rewarding investors that have generated in the past real 'change the game' inventions and that, as we can find, are delivering results still today. It is also true that studying the age distribution in a dynamic and not static way, the natural evolution for some Prolific investors will be to become Star.

4.1.2 Educational statistics

The aim of this section is trying to determine if the inventor's type and level of education and /or career can have an impact in becoming a Star Inventor. The first education level considered is the bachelor's degree resumed in *Table 4.2* Depending on the country of origin, it may last from three to five years. Generally, most part of the analyzed inventors have a bachelor's degree (BA), only 9% of them does not. The percentages are slightly different for the two samples, 93% of Star inventors have a bachelor while the percentage drops to 88% Prolific for the Prolific sample.

Table 4.2: Percentages of inventors with a Bachelor Degree

	% Star	% Prolific	All
Has BA	93%	88%	91%
Has not BA	7%	12%	9%

The second education level considered is the Master degree; data are resumed in *Table 4.3*.

Table 4.3: Percentages of inventors with a Master Degree

	% Star	% Prolific	All
Has MA	68%	60%	64%
Has not MA	32%	40%	36%

The percentage of inventors who have a Master Degree (MA) is obviously lower than that of the BA (only 64% of the inventors against 91%). Once again the percentage of Star is higher than the percentage of Prolific inventors (68%, 60% respectively). The distance between the Star and Prolific samples now grows from 5% of the Bachelor's Degree to 8% of Master Degree. The last educational variable taken into account is the international study experience. This is presented in *Table 4.4*.

Table 4.4: Percentages of inventors with an International study Experience

International Study Experience	% Star	% Prolific	All
Yes	18%	19%	19%
No	82%	81%	81%

The International study experience variable does not seem to be a characterizing element for the two samples. The percentages of inventors who studied out of their nation is almost the same for the two categories (18% for Star inventors and 19% for Prolific Inventors).

4.1.3 Career statistics

In this section, we want to study some statistics regarding an inventor's working experience. One can find information regarding the inventors' first activity (if he/she has started an Entrepreneurial activity) and the current activity (divided into two categories: company or academic profession).

Table 4.5: Percentages of inventors who started as entrepreneurs

First Experience as Entrepeneur	% Star	% Prolific	All
Yes	15%	9%	12%
No	85%	91%	88%

Referring to *Table 4.5* we can assert that the Inventors' attitude is not referred to the Entrepreneur one. Only 15% of Stars and 9% of Prolific started having their own job company/activity. This difference can be explained thinking that a Star inventor, after having patented his/her invention, decided to found their own business. On the other hand, as one can see in *Figure 4.8* and *Figure 4.9*, Prolific Inventors are Employed in private or public Companies for 98% and 33% have a University Job. Star have a percentage in an Academic/University Job of 68%, more than twice the percentage of Prolific and this is a clear distinguishing factor between the two populations. On the contrary, 30% of Star population has never worked in a company against 2% of Prolific Inventors.

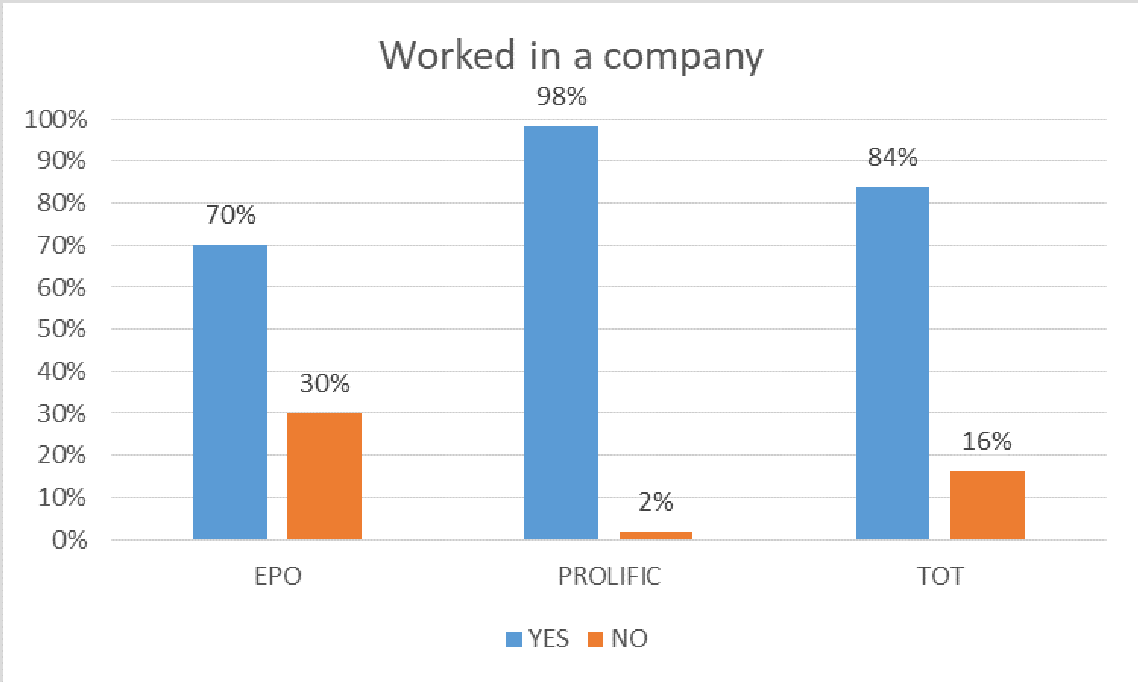

Figure 4.8: Inventors who worked in a company

4.2 Patent statistics

This section contains the core of this thesis. First of all the result of the extraction using the Derwent Innovation software will be briefly presented, then the result of some statistics will be described in order to allow the reader to understand the main patent characteristics that support our main distinction between Star and Prolific inventors. After the extractions, our sample of patents for Star and Prolific inventors is thus composed:

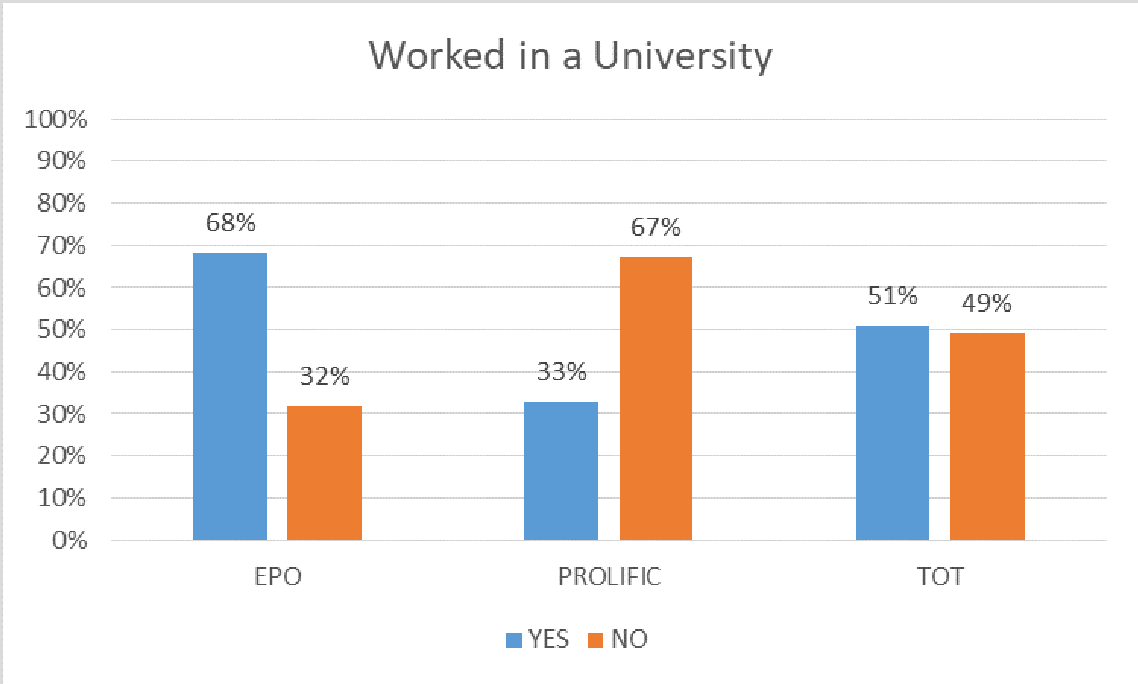

Figure 4.9: Inventors who worked in a university

- 12.301 records for Star inventors;
- 51.670 records for Prolific Inventors.

Then the following variables have been calculated for both Star and Prolific samples:

- Number of Patents;
- Number of Inventors;
- Number of Inventors;
- Number of Assignee;
- Number of Forward Citations Weighted;
- Number of Backward Citations - no patent;
- Number of Backward Citations – patent;

- Number of Backward Citations – total;
- Number of family members;
- Number of IPC Class.

For each of these variables, the average mean has been calculated and, in order to understand the dispersion of each of them, the standard deviation has been used. Then, by using a t-test, we tried to understand if the found values could be considered significantly different between the two analyzed samples. The null hypothesis that will be tested will regard the comparison between the two means; there follows the proper hypothesis definition:

$$H_0 : \mu_{STAR} = \mu_{PROLIFIC} \tag{4.1}$$

In order to test this hypothesis we selected an error of first kind of 5% (that is equivalent to a confidence interval of 95%). That means that, if the t-test gives a value smaller than 0.05 we can reject the null hypothesis by saying that our means are considerably different. *Table 4.6* below resumes the calculated variables.

Our basic distinction between the two samples of Star and Prolific inventors was based on the idea that the number of patents filed by each of them was significantly different; the assumption is now justified by the found data. The Prolific inventors have an average number of filed patents four times bigger than the Star sample (890,86 against 205,02). In addition, the standard deviation is significantly different; the Prolific sample has a std dev four times bigger than the Star sample. The difference between the two samples is finally proven by our t-test result. The obtained p-value for our test is 0,00054 that is smaller than 0,05; in this way we can reject the null hypothesis and we can consider the two samples' means as different. The different patent quality between the Prolific and Star sample was another standpoint of our analysis; the average of forward citations assesses this idea. Here an important

Table 4.6: Patents' statistics for Star and Prolific samples

Patent Statistics	All	Star	Prolific	t-test
Average number of patents	542,13	205,02	890,86	***
Std Dev Number of patents	1060,10	336,30	1395,11	
Average of inventors	3,90	3,43	4,39	***
Std Dev of inventors	1,82	1,48	2,00	
Average of assignee	1,52	1,55	1,50	
Std Dev of Assignee	0,50	0,64	0,28	
Average of weighted forward citations	0,32	0,38	0,25	***
Std Dev of weighted forward citations	0,29	0,36	0,16	
Average of backward citations - non patent	3,35	3,89	2,79	
Average of backward citations - patent	5,73	5,47	6,00	
Average of backward citations (non patent & patent)	9,08	9,36	8,79	
Std Dev of bkw cit (non patent & patent)	10,40	13,20	6,45	
Average of Number of family memebers	21,60	17,61	25,73	
Std Dev of family members	35,53	15,10	48,18	
Average of Number of IPC class	1,77	1,75	1,79	
Std Dev of IPC class	0,57	1,69	0,54	

insight must be done in order to understand why a weighted measure for forward citations has been used. Of course, the number of citations that a patent receives depend on the year of publication of each patent. In particular, older patents are more likely to have a higher number of forward citations. The use of a weighted measure wants to assess for this problem. For each patent, the number of years since its publication has been calculated and the number of forward citations has been divided by this value. The described equation is reported below:

$$Weighted\,Forward\,Citations = \frac{Number\,of\,Forward\,Citations}{2020 - Publication\,year} \tag{4.2}$$

We use 2020 as actual year because there are some patents that have been published during 2019, in this way we avoided the problem of the denominator of the fraction. The obtained result is really meaningful, even if the difference between the average number is not so high (0.38 for Star Inventors and 0.25 for Prolific Inventors), no doubt about the statistical significance of this result, the p-value obtained from the t-test allows us to reject the null hypothesis of equivalence between the two means. The Star inventors' patents can hence be considered of higher quality than that of Prolific inventors. What is interesting is the result obtained for the average number of backward citations. In this case, the null hypothesis regarding the conformity of means cannot be rejected but it is not important. The focus of the variable is linked to its hidden significance, the higher the number of backward citations to non-patent literature is, the nearest to the base science the invention is. This means that Star inventors' patents have more citations referring to the scientific literature and are nearest to the base science. This result is in line with the previous one referring to the working place of Star inventors. Given that the majority of Star inventors work in a University or research centers, their inventions are less linked to industrial applicability of the patent and report a higher number or reference to literature. On the other hand, the average number of reference to patents proves the same result.

Despite the result of our t-test, Prolific inventors' patents – as anticipated – are more subject to industrial application and report a slightly bigger number of backward citations referred to previous patents. The result is in line with the fact that Prolific inventors use to work more in companies than in academic environment. Despite the difference of the employment (University or Company) the average number of Assignee per patent is almost the same for both Star and Prolific samples. This can be interpreted in an interesting way. No or only a small number of Star patents would have an extraordinary success without an industrial application. On the other hand, it is important that an invention owns solid scientific bases in order to be planned. This may incentivize universities to collaborate with companies and vice versa in order to create patents bringing to an equal number of Assignee for the two samples. This last idea can be justified by also using the average number of inventors. Given that Prolific inventors work mainly in companies and their patents tend to have a real application, the number of figures needed to create an invention is higher than that of Stars inventors. In chapter 2 we introduced the fact that, in order to patent breakthrough inventions one must have deep knowledge on the subject he is studying; for star inventions, few figures with an extraordinary knowledge may be sufficient for an extraordinary patent. The deeper matter specialization of Star inventors can also be seen in the average number of IPC class per patent. Although the p-value is not significant, the average number of the two samples brings some differences. Every patent is filed, on average, in 1.75 IPC classes for Star inventors and 1.79 for Prolific inventors. The standard deviation, in this case, summarizes the main difference. Breakthrough inventions are those which manage to combine different knowledge; by comparing the two standard deviations we can see that Star Inventors' one is three times higher than Prolific inventors' one (1,69 against 0,54). The last interesting variable to be understood is the average number of family members. In this case, even if we cannot reject the null hypothesis of correspondence between the two values, a difference characterizes the two samples due to patent strategy and patent costs. To patent in different nations has a high cost impact, this may be significant for universities that can decide to patent only in few nations in order to avoid higher costs. Prolific inventors on the other hand, by working in companies, have access to higher funds and can decide to cover a higher number of nations with their patents. This can be the reason of the difference between the two samples. There follows a final application of the found data.

In chapter 2, we introduced Oettl's new taxonomy for star inventors: he defined four sub-groups of inventors depending on their collaboration and productivity. Here we adapted his classification with our variables; we used the average number of weighted forward citations as proxy for patents' quality and the average number of inventors as proxy of helpfulness. We then categorized our samples of inventors into one of the four Oettl's sub-groups: Non-Star, Lone wolf, Maven, All-star. The final inventors' classification confirmed our previous results: for Prolific inventors, the most populated category is the Maven one (inventors with a high helpfulness index and an average patent quality) while the less popular was the Lone Wolf. Different results have been obtained for Star inventors whose distribution appear uniform in the four categories.

Chapter 5

Regression Analysis

5.1 The adopted methodology

In order to understand if the distinction between Star and Prolific inventors can be observed through the information contained in the patents they filed, a regression analysis has been used. In particular, we chose a multivariate logit regression because it studies the relationship between a binary dependent variable (being or not a Star Inventor) and one or more nominal, ordinal, interval or ratio independent variables. Our dependent variable will assume the value 1 if the specific inventor has won or participated as a finalist at the EPO Award, 0 otherwise. The independent variables are a subgroup of those aforementioned, they are reported below:

- *GenderDummy*: it assumes 1 if the inventor is a female, 0 if he is a male;
- *UniDummy*: it assumes 1 if the inventor has worked in a university or research center, 0 if not;

While the first two variables refer to biographical and educational experiences, the following ones refer to the main information that we found from the patents' extraction:

- *AvgPatents*: average number of patents;

- *AvgInventors*: average number of inventors;
- *AvgAssignee*: average number of assignee;
- *AvgFwdCitWeighted*: average number of weighted forward citations;
- *AvgBwdCitnonPat*: average number of backward citations not referred to patents;
- *AvgFamilyMemb*: average number of family members for the specific patent;
- *AvgIPCCount*: average number of IPC category where the patent has been filed.

The regression analysis, implemented with STATA software, will give an idea of which of the independent variables can be used in order to understand the specific state taken by the dependent variable.

5.2 The logit regression analysis

In order to understand if the following regression analysis could be negatively influenced by any correlation between the analyzed variables, a correlation test has been done. The STATA outputs reported below show the most significant correlations found between some variables. First, a strong correlation has been found between the variables GenderDummy and AvgAssignee such that we decided to repeat our models without the control variable GenderDummy (Model 5 to Model 8). The correlation between these two variables is shown in *Figure 5.1*.

```
. pwcorr GenderDummy AvgAssignee
```

	Gender~e	AvgAss~e
GenderDumm~e	1.0000	
AvgAssignee	0.4133	1.0000

Figure 5.1: Correlation between GenderDummy and AvgAssignee

A second strong correlation has been found between the variables AvgAssignee and AvgInventors as reported in *Figure 5.2*.

```
. pwcorr AvgAssignee AvgInventors
```

	AvgAss~e	AvgInv~s
AvgAssignee	1.0000	
AvgInventors	0.4779	1.0000

Figure 5.2: Correlation between AvgAssignee and AvgInventors

As regards the link between these two variables, we decided to introduce two more regression models where we alternated the variables AvgAssignee and AvgInventors in order to understand if this correlation negatively influenced our analysis. *Figure 5.3* shows the results of the regression analyses. The aim of the regression analysis is to test if some of the independent variables may foresee the fact of being a star or prolific inventor. In order to understand if the relationship between the dependent variable and the variables of interest is Non-Significant, Significant or Very Significant, more than one model has been tested. In particular, eight models have been implemented and, if the same variable results Significant or Very-Significant in every model, it can be considered Robust. The models differ for the variables used; the following list presents each of them:

- **Model 1**: it highlights the results of a regression analysis between the EPO depend variable and all the variables under the inventors' direct control (thus we have excluded from this regression the average number of forward citations);
- **Model 2**: this model reports the result of the full regression analysis, all the control variables and the significant variables have been used in order to understand the relationship with being or not a Star inventor;
- **Model 3**: according to the result of the correlation test, the variables AvgInventors and AvgAssignee are correlated. In the current model we have excluded

AvgInventors in order to avoid dependence errors;

- **Model 4**: with respect to the Model 3, in the current model we included the AvgInventors variable and we excluded the AvgAssignee one;

The following models, compared to the previous one, exclude the GenderDummy variable. We decided to do so because the variable seems to be significant with respect to AvgAssignee. The previous tests have been repeated with the exclusion of this variable.

- **Model 5**: it reports the result of a regression analysis including the variable under inventors' control (it excludes GenderDummy as aforementioned, and AvgFwdCitWeighted);

- **Model 6**: it repeats the Model 2 regression analysis except for the GenderDummy variable;

- **Model 7**: it repeats the Model 3 regression analysis except for the GenderDummy variable;

- **Model 8**: it repeats the Model 4 regression analysis except for the GenderDummy variable;

The *Figure 5.3* reports the results of the regression analyses.

The purpose of this study is to define if there are links between becoming a Star Inventor and having some particular features in the patents' characteristics. Firstly, even if from Model 5 to Model 8, the variable *GenderDummy* has been excluded, we can see that every model is robust and the exclusion or inclusion of this current variable does not change the results of the regression analysis; for this reason we will discuss the results of Model 1 to Model 4 because they include every significant information. Model 1 and Model 2 are our starting models, Model 3 and Model 4 represent two robustness check models. Considering the results of these models, the

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Model1	Model2	Model3	Model4	Model5	Model6	Model7	Model8
GenderDummy	2.278	2.837*	1.915*	2.857*				
	(1.700)	(1.644)	(1.147)	(1.480)				
UniDummy	1.197**	1.103**	0.907*	1.174**	1.167**	1.037*	0.870*	1.155**
	(0.516)	(0.546)	(0.493)	(0.542)	(0.512)	(0.534)	(0.487)	(0.524)
AvgPatents	-0.003***	-0.003***	-0.003***	-0.003***	-0.003***	-0.003***	-0.004***	-0.003***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
AvgInventors	-0.607***	-0.646***		-0.553***	-0.606***	-0.658***		-0.488***
	(0.196)	(0.209)		(0.178)	(0.210)	(0.230)		(0.184)
AvgAssignee	0.317	0.702	-0.386		0.561	0.979	-0.125	
	(0.641)	(0.701)	(0.535)		(0.614)	(0.678)	(0.499)	
AvgFwdCitWeighted		3.666***	3.598***	3.499***		3.128**	3.114**	2.823**
		(1.319)	(1.296)	(1.308)		(1.216)	(1.215)	(1.194)
AvgBwdCitnonPat	0.053	-0.026	-0.010	-0.006	0.041	-0.016	0.001	0.010
	(0.061)	(0.075)	(0.060)	(0.071)	(0.056)	(0.059)	(0.051)	(0.054)
AvgFamilyMemb	-0.001	-0.005	-0.004	-0.004	0.004	0.003	0.001	0.004
	(0.009)	(0.009)	(0.010)	(0.009)	(0.009)	(0.009)	(0.010)	(0.009)
AvgIPCCount	0.205	0.860	0.537	0.807	0.176	0.715	0.399	0.625
	(0.519)	(0.602)	(0.528)	(0.591)	(0.513)	(0.584)	(0.512)	(0.567)
Constant	1.670	-0.670	-0.644	0.124	1.444	-0.547	-0.566	0.504
	(1.160)	(1.448)	(1.249)	(1.187)	(1.118)	(1.388)	(1.212)	(1.143)
Observations	118	118	118	118	118	118	118	118
Loglike	-54.72	-49.41	-55.93	-49.95	-56.05	-51.71	-57.67	-52.90
Pseudo R2	0.3309	0.3958	0.3161	0.3892	0.3146	0.3676	0.2947	0.3530

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.10

Figure 5.3: Regression analyses

first significant variable that seems to be important in relation to be a Star inventor is the *UniDummy*. It remains significant in every regression test done, therefore it can be considered robust: it seems that those who worked in a university have more probability to become a Star Inventor. The same result can be observed for the *AvgPatents*: being a Star Inventor is negatively correlated to the number of patents filed. This is an important result because we used this variable (along with patent quality) in order to discriminate between Star and Prolific inventors. Another interesting result is the negative correlation between the dependent variable and the number of inventors per patent (*AvgInventors*). It seems that Prolific Inventors are more collaborative than Star Inventors. The same result was analyzed in Chapter 4 where the t-test done highlighted a significant difference between the average numbers of inventors of the two groups analyzed. On the contrary, there is no significant relation between having won/participated to the EPO Award and the average number of assignee per patent (*AvgAssignee*). Regarding the two most important variables measuring patent quality, we found a Very-Significant and thus robust correlation between being a Star Inventor and the average number of citations that a patent receives (*AvgFwdCitWeighted*). As explained in Chapter 4, the *AvgFwdCitWeighted* variable is not influenced by the age of an invention so this represents an important and robust result. The second variable used in order to discriminate the quality of an invention is the average number of family members; no confirmation of this assumption is obtained from the regression analysis. Such variable does not foresee being a Star or a Prolific inventor. An interesting result is also given by the *AvgBwdCitnonPat* variable. In spite of what has been said in Chapter 4 regarding the affinity of Star Inventors' patent to the base science, no confirmation of this assumption has been found. Being or not a Star Inventor does not appear to be influenced by the average number of citations to the base science.

5.3 General conclusion on the regression analysis

Thanks to iterations of the regression analyses and the results anticipated by the statistical analysis (Chapter 4), the two samples of inventors can be distinguished by the results found. Regarding our control variable, it is more probable to become a Star inventor if a subject has worked in a university or a research center. Regarding our variable of interest, what is important is the positive and strong correlation between the average number of forward citations and being a Star inventor. Moreover, what is equally important is the negative correlation between the dependent variable and the average number of patents filed by each inventor. In addition, we can say that Prolific Inventors are, generally, more collaborative than Star Inventors. The variables counting for the family members of each patent and the IPC category where a patent is filed, cannot be used to foresee if an Inventor can become a Star.

Chapter 6

Conclusion

This final chapter reports the main conclusions that can be drawn from our analysis. It also tries to anticipate the main criticism one can make to our methodology first and then to our model; then we try to answer it.

6.1 Final conclusions

The final aim of this thesis was to understand the differences between the two groups of Star and Prolifc inventors by comparing the features of the patents filed by each of them and to derive implications in terms of impact and quality of the inventors' output as regards the total amount of their production.

Starting from a group of inventors, 60 Star inventors and 58 Prolific inventors have been randomly selected. The Star inventors are those who have won or participated as a finalist at the EPO Award and they are opposed to the Prolific Inventors. For each of them, his/her patent portfolio has been extracted using Derwent Innovation software: information about 63.971 patents have been collected and then analyzed. In order to export the patents' portfolio of each inventor, a query has been elaborated. Someone can argue that the dimensions of some patents' portfolio can be overestimated due to the fact that some inventors may be filed in the patent application only because of the role covered in the company. This problem is not under our control so we assume that the filed inventors always give an important contribution in managerial, knowledge or operative terms so that they deserve to be considered part of the inventors' team. Another criticism that someone can make to our model was linked to the use of simple forward citations that are obviously influenced by the patent application year. This issue has been solved by using a weighted measure for forward citations that standardizes each observation. Then two samples of inventors have been analyzed: using *Excel* the main inventors' biographical and educational statistics have been studied, these variables found were important because they have been used as control variables of the regression analysis. Then the focus passed on the main patents' information. First, some statistics have been calculated and then, thanks to the logit regression analysis, some conclusions have been drawn. In particular, we can now say that those inventors who worked in a university or research center have a higher probability to become Star inventors. However, it is important to specify, in order that an invention obtains the deserved success, a focus on what really 'matter to people' must be present. Ideally, the inventors who have the capability to bridge business requirements and research capability have 'what needed' to become a Star. Furthermore, the average number of forward citations per patent and the number of patents in each inventors' portfolio have highlighted another important result. The significance and robustness of both these two variables is important because it firstly confirms our distinction between these two categories of inventors and allows us to say that Star Inventors, on average, produce fewer patents than Prolific Inventors. Secondly, an important result about the production quality has been found: Star inventors' output is, on average, more valuable than prolific inventors' one. These results are of fundamental importance and, as far as we know, no other works have already achieved them.

Bibliography

[1] Joseph A Schumpeter. *Capitalism, socialism and democracy.* routledge, 2010.

[2] Theodore Levitt et al. *Exploit the product life cycle.* Vol. 43. Graduate School of Business Administration, Harvard University, 1965.

[3] René Mauborgne and W Chan Kim. "Blue Ocean Strategy". In: *Harvard Business School Publishing Corporation. Boston, Massachusetts* (2005).

[4] Gautam Ahuja and Curba Morris Lampert. "Entrepreneurship in the large corporation: A longitudinal study of how established firms create breakthrough inventions". In: *Strategic management journal* 22.6-7 (2001), pp. 521–543.

[5] Lynne G Zucker and Michael R Darby. "Star scientists and institutional transformation: Patterns of invention and innovation in the formation of the biotechnology industry". In: *Proceedings of the National Academy of Sciences* 93.23 (1996), pp. 12709–12716.

[6] Jorge Niosi and Johanne Queenton. "Knowledge capital in biotechnology industry: impacts on Canadian firm performance". In: *International Journal of Knowledge-Based Development* 1.1-2 (2010), pp. 136–151.

[7] Xue Han and Jorge Niosi. "Star scientists in PV technology and the limits of academic entrepreneurship". In: *The Revolution in Energy Technology.* Edward Elgar Publishing, 2018.

[8] Jan Hohberger. “Does it pay to stand on the shoulders of giants? An analysis of the inventions of star inventors in the biotechnology sector”. In: *Research Policy* 45.3 (2016), pp. 682–698.

[9] Christian Le Bas, Riad Bouklia-Hassane, and Alexandre Cabagnols. “Prolific inventors: Who are they and where do they locate? Evidence from a five countries us patenting data set”. In: *Evidence from a Five Countries US Patenting Data Set (June 16, 2010). International Centre for Economic Research Working Paper* 14 (2010).

[10] L Zucker, MR Darby, and MB Brewer. ““Intellectual human capital and the birth of US biotechnology enterprises”American Economic Review 88 290ˆ 306 Conditions of use”. In: *This article may be downloaded from the E&P website for personal research by members of subscribing organisations. This PDF may not be placed on any website (or other online distribution system) without permission of the publisher* (1998).

[11] Lydia Christina Reichensperger. “Three essays on inventors, inventions, and innovation”. In: (2018).

[12] Christiane Goetze. “An empirical enquiry into co-patent networks and their stars: The case of cardiac pacemaker technology”. In: *Technovation* 30.7-8 (2010), pp. 436–446.

[13] Zhenzhong Ma and Yender Lee. “Patent application and technological collaboration in inventive activities: 1980–2005”. In: *Technovation* 28.6 (2008), pp. 379–390.

[14] Benjamin F Jones. “The burden of knowledge and the “death of the renaissance man”: Is innovation getting harder?” In: *The Review of Economic Studies* 76.1 (2009), pp. 283–317.

[15] Alexander Oettl. “Reconceptualizing stars: Scientist helpfulness and peer performance”. In: *Management Science* 58.6 (2012), pp. 1122–1140.

[16] Tehmina Amjad et al. “Standing on the shoulders of giants”. In: *Journal of Informetrics* 11.1 (2017), pp. 307–323.

[17] Paola Giuri et al. “Inventors and invention processes in Europe: Results from the PatVal-EU survey”. In: *Research policy* 36.8 (2007), pp. 1107–1127.

[18] Sara Ferragatti. “Star and Prolific inventors: Empirical analysis of differences in education and career”. PhD thesis. Politecnico di Torino, 2019.

[20] Richard C Levin et al. “Appropriating the returns from industrial research and development”. In: *Brookings papers on economic activity* 1987.3 (1987), pp. 783–831.

[23] Roberto Fontana et al. “Reassessing patent propensity: Evidence from a dataset of R&D awards, 1977–2004”. In: *Research Policy* 42.10 (2013), pp. 1780–1792.

[24] Bronwyn H Hall, Adam Jaffe, and Manuel Trajtenberg. “Market value and patent citations”. In: *RAND Journal of economics* (2005), pp. 16–38.

[25] James Bessen. “The value of US patents by owner and patent characteristics”. In: *Research Policy* 37.5 (2008), pp. 932–945.

[26] Dietmar Harhoff, Frederic M Scherer, and Katrin Vopel. “Citations, family size, opposition and the value of patent rights”. In: *Research policy* 32.8 (2003), pp. 1343–1363.

[27] Jean O Lanjouw and Mark Schankerman. *Stylized facts of patent litigation: Value, scope.* Tech. rep. and ownership. Working paper w6297, National Bureau of Economic Research ..., 1997.

Sitography

[19] URL: https://www.epo.org/learning-events/european-inventor/about.html.

[21] URL: https://www.epo.org/law-practice/legal-texts/html/epc/2016/e/ar52.html.

[22] URL: https://www.wipo.int/treaties/en/text.jsp?file_id=288514#P83_6610.

[28] URL: https://www.epo.org/searching-for-patents/helpful-resources/first-time-here/patent-families.html.

[29] URL: https://www.wipo.int/classifications/ipc/en/.

[30] URL: https://data.oecd.org/rd/gross-domestic-spending-on-r-d.htm.

[31] URL: miur%5C%20distribuzione%5C%20ricercatori.pdf.

[32] URL: https://www.bbc.com/news/technology-49843990.

[33] URL: https://data.oecd.org/rd/gross-domestic-spending-on-r-d.htm.

[34] URL: https://www.ilsole24ore.com/art/spesa-rd-l-italia-investe-quarto-germania-AEq0BdLD?refresh_ce=1.