ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Bachelor Thesis Strategy Economics

Do innovative companies stimulate new innovation?

A patent data analysis

Name: Tim Blaauw

Student Number: 481702

Supervisor: Ajay Bhaskarabhatla

Second Assessor: Niels Rietveld

Date final version: 19-07-2021

Abstract

This bachelor thesis investigates the relationship between the number and quality of patents applied for by incumbent firms and the number of patents applied for by entrants. The data used for this research is panel dataset of patents applied for in the United States between 1976 and 2006 at the patents class level. This research is motivated by the paper from Trajtenberg, Henderson and Jaffe (1997) in which they introduce the generality and originality measures of basicness to assess the quality of patents. This thesis connects these concepts to the causal effect patents from incumbent firms have on patents from entrants and whether the former encourages the latter. This research has significant policy implications and adds to scientific research regarding patenting. The key findings are that the number of patents from incumbents positively influences the number of patents by entrants in a given year and that forward citations of patents in general have a negative impact, while backward citations have a positive effect. Moreover, generality and originality do not have direct significant effects on the number of patents from entrants. Interestingly, generality becomes significant when interacting with the number of entrants, while lacking significance in all other models. Lastly, the outcomes of the thesis suggest that all the effects discussed differ per patent category.

KEYWORDS: *Innovation, Patents, Generality, Originality*

Table of Contents

Introduction

Enhancing firm productivity is an essential component for the survival of a firm. In the current fast-changing society competitive advantages can practically evaporate overnight if competitors produce faster, set lower prices or guarantee a higher quality standard. Modern growth theory states that productivity growth consists of three determinants: human capital, physical capital and, most important, innovation (Rao et al., 2001). Innovation is a key driver for firm performance and a measure for technological advancement at the industry and firm level. Innovation can therefore be defined as the process of introducing new ideas to a firm which result in increased firm performance. Therefore, innovation is necessary to increase firm performance.

Innovation is a complex process that covers a broad range of activities that can differ across firms, making it a difficult phenomenon to measure. As Simon Kuznets (1962) mentioned*: "perhaps the greatest obstacle to understanding the role of innovation in economic processes has been the lack of meaningful measures of innovative inputs and outputs"* Luckily, more recent research has addressed this problem. For example, Rogers (1998) divides the measurement into output and input measures, like Kuznets suggested. Output measures are the outputs of innovative activity, such as profits, share performance, number of new or improved products introduced and intellectual property statistics. Examples of input measures are research and development expenditures, acquisition of technology from others, marketing expenditures for new products and intangible assets on the balance sheet such as increased goodwill. These statistics are all possible ways to quantify the level of innovation in a firm or industry.

However, patents are most often used as a proxy indicator for innovation (Kleinknecht, 2000). A patent can be seen as the exclusive right to the inventor that created something new and useful and can be used for commercial gain. Therefore, it is an output measure as defined by Rogers (1998). Past research has discovered that firms that obtain more patents record higher levels of innovation, meaning that having more patents is a good indicator of the innovativeness of a firm (Mansfield, 1986).

An interesting aspect of innovation is the environment in which this happens. Is encouraging patenting an effective way to boost innovation? Does the government have access to more effective tools to achieve this? Moser (2013), for example, has done research on the role the patent system plays in encouraging innovation. She shows historical evidence that countries with and without strong patent laws still display the same level of innovation. Therefore, she questions the effect patenting has on innovation, stating that patents that are too broad can even prevent an industry from moving forward in terms of innovation. Still, the fact is that companies must operate in a world where innovations can be patented and they can choose themselves whether or not to trademark their new idea or keep it a company secret.

1.1 Research Question

This thesis focuses on what happens if companies eventually choose to patent their innovation and the consequences of this decision for other firms. Governments can impose regulations that boost patent applications or encourage innovation in general, but actions like this lead to a manufactured innovation environment. The natural environment created by universities and companies themselves should have a positive influence on the number of patents that are applied for, while for example Moser (2013) showed that it is debatable whether the manufactured innovation environment positively influences innovation. This leads to the question whether companies are inspired by other patenting companies or feel the need that they cannot be left behind without further innovation. Does the decision to patent an innovation have positive spill over effects for other companies? If experienced companies that have applied for several patents in previous years apply for a lot of patents in a certain year, does that encourage other companies to apply for their first patent? This leads to the research question of this thesis:

How do quality and number of patents from incumbent firms impact the amount of firms applying for their first patent?

In this research incumbent firms are identified as firms that have applied for patents in previous years and entrants are firms applying for the first patent in company history.

This research is relevant in several different ways. First of all, from a policy point of view; if the relationship between the quality and number of patents by incumbent firms and the number of patents applied for by entrants is positive, this would mean that policy choices regarding the patent system could be effective after all. More patents by incumbents would lead to more firms applying for their first patent, which increases the general level of innovation in a country.

Secondly, the outcome of this research is relevant from a firm point of view as well. If the relationship is positive, that would mean that by patenting a new idea, they would be kickstarting other firms to patent their innovations as well. This has positive consequences, namely that applying for a patent means that a new innovation is introduced to the world instead of kept secret. This can inspire incumbent firms for new innovations as well, creating a positive cycle of innovation inspiration. It can also have negative consequences. Incumbents that show high levels of patenting encourage competitors to innovate as well by showing the world their innovations via their patents. This may deteriorate an incumbents relative competitive advantage.

2. Theoretical framework

This chapter lays out the theoretical underpinnings of patents and their influence on innovation. Firstly, the relationship between innovation and patents will be discussed. Secondly, an overview of previous research on the impact of the number and quality of patents will be displayed. The theoretical framework then moves on by summarizing the paper this thesis builds upon. Lastly, the hypotheses of this thesis will be presented.

2.1 Patents and innovation

This thesis discusses a relative narrow aspect of patenting, namely how patenting by incumbents influences patenting by new entrants, but its scope is much wider. The implications of this research are appropriate in the context of innovation as a whole. Conclusions from this research can have policy implications that are applicable to enhancing innovation in an industry and assessing whether patents are an adequate tool for governments when enhancing innovation. Therefore, it is important to summarize the key findings of previous research regarding the relationship between patents and innovation and if more patents lead to more innovation.

The traditional view of this relationship has emphasized the importance of patents as a primary driver of innovation. However, new datasets and advancements in empirical methodology have started to challenge this traditional view, according to Moser (2016). The reasoning behind the traditional view is that without intellectual property rights to protect new ideas, innovation would not or barely exist due to free-rider problems. Without legal protection, every idea can be used by anybody, without compensating the original inventor. As a result, the inventor loses incentive to invest time and resources in research and development of a new idea, because the inventor does not have the power to prevent others from using the idea when it is publicised and can therefore not economically benefit from the new idea. A patent solves this issue. It forms a legal basis for the inventor to prevent anybody from using the new idea without appropriate compensation. Moser (2016) states in her research that these traditional predictions are somewhat ambiguous and are supported by little empirical evidence. Using empirical evidence from world fares in the 19th century, she concludes that many important innovations occur outside of the patent system and that patent laws are not necessary for innovation. Countries that did not have patent protection still contributed heavily to the world fare, where traditionally new inventions from each participating country are displayed for the world to see. A possible explanation laid out by Moser is that it takes time and money to patent an invention, which is saved when the invention is simply kept secret. However, empirical evidence shows that in countries with lower patent application costs innovation is not higher than when these costs are relatively high. The actual explanation is that enforcement costs of a patent are extremely high due to risk of litigation when the patent is applied for. Applying for a patent means that the invention becomes publicly known, opening up the possibility for litigation.

The question arises why certain companies prefer patenting and others prefer secrecy. Cohen, Nelson and Wash (2000) found via a questionnaire that firms protect profits by using patents, secrecy, lead time advantages and complementary marketing and manufacturing capabilities. The manufacturing firms that were a part of the survey reported that patenting was the least used mechanism to protect profits. They find that secrecy now appears to be used much more than patenting across most industries than previously. Still patenting is not a tool of the past, there are simply other motives than directly profiting from a patented idea. Motives can be the prevention of rivals from patenting related inventions (patent blocking), the use of patents in negotiations and the prevention of lawsuits. In industries where these motives are not applicable, firms prefer secrecy above patenting. Industries that prefer secrecy are telecommunications and semiconductors, while chemical and pharmaceutical industries prefer patenting, because the alternative motives are more applicable.

In contradiction to Moser (2016), Lamoreaux & Sokoloff (1999) are more supportive of the role patents play in innovation, which is the more traditional view on the topic. In their research, they accredit the sharp acceleration of innovative activity in the United States in the nineteenth century to the large economic growth in that same period. The evidence for this claim is originated in patent records, forming a proxy for innovation that is coupled to accelerated economic growth. The main conclusion is twofold. Firstly, they celebrate the patent system for its crucial role in accelerating innovation, because of the granting of temporary monopoly rights for the use of their discoveries. Secondly, specification in US patent law of tradable assets in technology also benefits innovation.

However, the traditional view is being debunked lately by scientists such as Moser (2016), Moser (2013), Pakes and Griliches (1980), saying: “patents are a flawed measure (of innovative output) particularly since not all new innovations are patented and since patents differ greatly in their economic impact” and Griliches (1990). Even early doubt on the positive relationship was expressed by Acs and Audretsch (1988). They use research and development (R&D) expenditures to link innovation with patented inventions. While first finding that the total number of innovations is closely linked to industry R&D expenditures and patented inventions, they also find that the relationship between R&D and innovation is somewhat different from that between R&D and patents. This is one of the earliest indications that patenting may not directly lead to innovation.

These conclusions from previous research indicate how ambiguous the effect patents have on innovation truly is. That ambiguity returns when diving deeper into two aspects of patenting: patent pools and compulsory licensing.

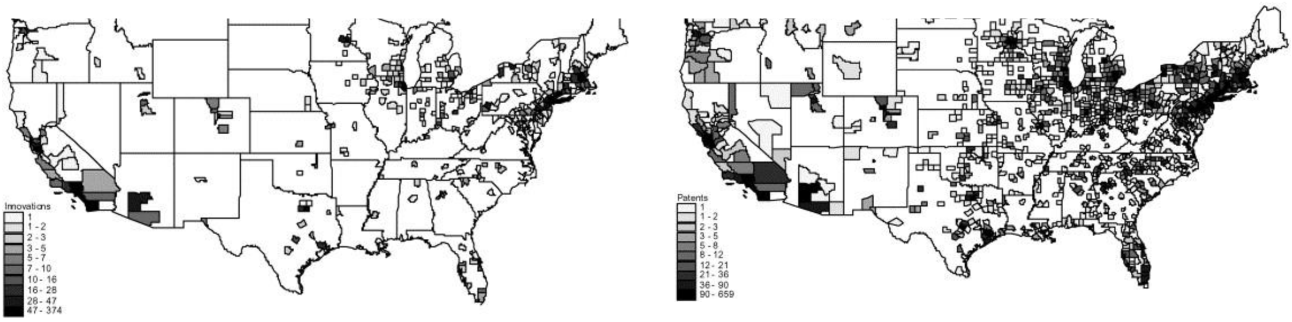
Lampe and Moser (2010) researched the influence of patent pools on innovation. A patent pool is when several companies agree to use a set of patents in the same industry as if they were jointly and equally owned by all members of the pool and license these patents as a package to other firms. The benefit is that the patent pool blocks a larger segment of competitors from coming up with innovations that make your innovation obsolete. They use evidence from the nineteenth century sewing machine industry, because this was the first patent pool in existence. The idea is that pooling technologies will prevent excessive litigation by close competitors, thus making it more attractive to innovate. In the example of the sewing machine industry, crippling litigation was ended by the patent pool, so in theory patent pools should encourage innovation. However, in reality pool members patented less in the years the pool existed and non-pool members patented more than usual in those years. Therefore, patent pools do not encourage innovation.

On the other hand, Moser & Voena (2012) find that compulsory licensing does have a positive effect on innovation. Compulsory licensing allows firms in developing countries to produce foreign inventions without the consent of foreign patent owners or risk of litigation. This is used to recreate life-saving drugs and technology to combat climate change, for example. The paper uses the 'Trading with the Enemy Act' in the Second World War, where US firms were allowed to violate German patents if they contributed to the war effort. This means that German patents were compulsory licensed. The result was that US domestic firms that licensed more German patents, went on to apply for more patents themselves. This shows that compulsory licensing has a positive effect on domestic invention, but this effect is lagged with almost a decade.

In conclusion, patents may not be the perfect encouragement for innovation, but research has found a different relationship between innovation and patents. Acs and Audretsch (1989) conclude that patents provide a fairly reliable measure of innovation. Hall, Jaffe and Trajtenberg (2001) say that patents do not measure economic value of new technologies, but that patents still are a good indicator of technology creation. Also, Acs, Anselin & Varga (2002) provide a thorough empirical regression analysis that indicates that patents provide a fairly reliable measure of innovation. They also prove this visibly, using the two maps of the continental US in figure 1. The map on the left shows the spatial distribution of innovations and the map on the right shows the spatial distribution of patents, both in 1982. The regression and the maps both indicate that patents are a good measure of innovation, even though it is ambiguous whether patents encourage innovation.

Figure 1

Distribution of innovations (left) and patents (right) in the US in 1982



Source: Acs, Anselin & Varga (2002).

2.2 The number and quality of patents and their relationship with innovation

This thesis investigates whether the number and quality of patents applied for by incumbents influences the number of patents applied for by entrants. Before empirically researching this relationship, it is important to survey the literature regarding the impact the number and quality of patents on the innovation environment.

The most commonly used quality measure for patents is number of citations. Just as in academic literature, when an inventor applies for a patent, he or she must mention the previous patents that the new patent builds upon. This is a backward citation. Forward citations are collected in the years after the patent has been approved. If a new patent references an original patent, the original patent receives a forward citation. This means that the technological advancement in the earlier patent has been built upon in the later innovation. There is a circular relationship between the two citation types. For instance, if patent 1 contains patent 2 as a forward citation, then patent 2 contains patent 1 as a backward citation. Katila (2000) found that citations are a good proxy for patent quality.

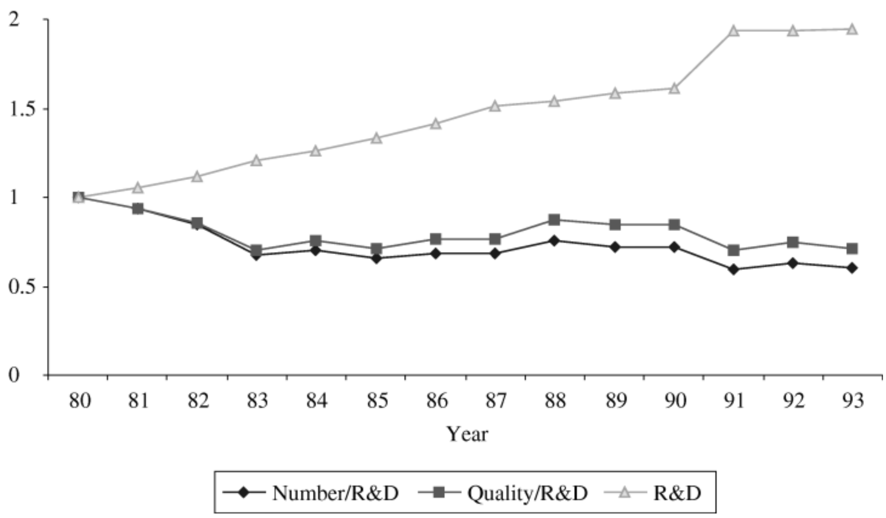
Sterzi (2013) has done research on whether ownership influences the quality of a patent. He uses the number of citations as a measure of quality. He found that academic patents owned by private firms receive more citations in the first years after the filing date than those owned by universities or other public research organizations. This difference decapitates when considering a longer time frame. Also, change of ownership is an indicator of patent quality. Academic patents owned by companies but originally assigned to universities or other public research organizations show a noticeably higher quality. This research indicates that ownership of a patent is a good way to assess its quality.

The quality of a patent also influences innovation in a sector. Tseng and Wu (2006) research whether innovation quality influences firm performance in the US automobile sector. They start off with identifying five indicators of innovation quality, namely patent count, relative citation ratio, citation-weighted patents, science linkage and scope of innovations. Using these innovation quality indicators, they discover that three of the five indicators, patent count, citation-weighted patents and scope of innovations, positively and significantly affect firm profits. This means that the quality of patents influences firm performance.

Lanjouw and Schankerman (2004) find that quality of patents owned by firms has a positive effect on the stock market value of firms. They also look at research productivity, measured by the ratio of patents to R&D expenditures. In the 40 years prior to the paper, this ratio has fallen sharply and they try to find an explanation for this. By 1990 the number of patents produced per US scientists and engineers had fallen to just 55% compared to its 1970 level. Figure 2 presents the time paths of R&D expenditures, unadjusted research productivity (the number of patents divided by R&D), and a patent quality-adjusted measure of research productivity for the mechanical industry from 1980 until 1993. As the graph shows, an increase in the quality of patents accounts for a sizable share of declines in research productivity when they are observed at the sector level. They conclude that quality of patents, does not appear to have a strong impact on research productivity at the firm level. It appears that more patents and higher R&D expenditures do not always lead to higher quality patents. The combination of R&D expenditures and quality of patents is therefore not an adequate measure for this thesis.

Figure 2

The time paths of R&D expenditures, unadjusted research productivity (the number of patents divided by R&D and a patent quality-adjusted measure of research productivity for the mechanical industry from 1980–93



Source: Lanjouw and Schankerman (2004).

2.3 Originality and generality

The most relevant measures of the quality of a patent for this thesis are originality and generality. These two measures were introduced by Trajtenberg, Henderson & Jaffe in 1997. They were chosen for this thesis because of availability in the data and because they form a variable that portrays backward and forward citations. This thesis builds upon their work.

Trajtenberg, Henderson & Jaffe (1997) research whether universities perform more basic research than corporations. They do so by looking at the 'basicness' of patents. This is useful because many widely used measures until that time, such as simple patent counts or counts of expert identified innovations, are limited in the fact that they cannot account for the heterogeneity of research projects and outcomes that characterizes the R&D process. They also look at the relationship between basicness and appropriability. Basicness refers to originality and generality of a patent and appropriability refers to if an inventor can make money off of the patent. Appropriability is a problem when making R&D investment decisions and Arrow (1962) stated that appropriability problems get larger when patents are more basic. For this thesis basicness measures are important. They are based on citations, because that provides evidence of the links between an innovation and its technological "antecedents" and "descendants". In their paper they construct a set of measures to capture key aspects of basicness and appropriability and test whether their measures capture basicness and appropriability.

Thus, research is regarded as basic if it "focuses on scientific rather than on technological questions; if it seeks to elucidate general laws rather than solving particular technical problems; and if it addresses old puzzles with original methods" (Trajtenberg, Henderson & Jaffe, 1997). Research *outcomes* are seen as basic if they have a major impact upon a given field or significant impact across a broad range of fields. These outcomes must then be fundamental to much later work and are often referred to and relied upon by scientists in the same or other fields. Seminal research is therefore basic.

The measures they create to capture basicness and appropriability can be divided into two categories. The first set of measures is backward-looking. These measures use the relationship between the patent and the knowledge that preceded it. They are relevant for assessing whether research is basic or appropriable. The second set of measures is forward-looking. They use the relationship between the patent and the technological developments that build upon it. They are relevant for assessing whether research outcomes are basic or appropriable.

Figure 3

Forward looking measures

1. Basicness Measures		
1.1 IMPORTF	Number of citing patents, including second generation cites	$\boldsymbol{IMPORTF_i = NCITING_i + \lambda \sum_{j=1}^{nciting} NCITING_{i+1j}}$
1.2 GENERAL	Herfindahl index on technological classes of citing patents	$\boldsymbol{GENERAL_i = 1 - \sum_{k=1}^{N_i} \left(\frac{NCITING_{ik}}{NCITING_i}\right)^2}$
2. Distance Measures		
2.1 TECHF	Distance in technology space	$TECHF_i = \sum_{j=1}^{nciting} \frac{TECH_j}{NCITING_i}$
2.2 TIMEF	Average citation lag	$TIMEF_i = \sum_{j=1}^{nciting} \frac{LAG_j}{NCITING_i}$
3. Appropriability **3.1 PSELF**	Number of self-citations	-

Source: Trajtenberg, Henderson & Jaffe (1997).

In figure 3 the forward-looking measures they created to test whether patents are basic or appropriable are displayed. The most important forward-looking measure for this thesis is generality. This is defined as: “the extent to which the follow-up technical advances are spread across multiple technological fields, computed with the Herfindahl-index” (Trajtenberg, Henderson & Jaffe, 1997). The other measures describe the amount of citations in later work (*IMPORTF*), distance in technological space and time (more basic innovations take longer to be cited by others) and ownership (*PSELF*).

The backward-looking measures are the same as the forward-looking measures, only instead of using *NCITING* in formulas, which denotes the number of patents citing the original patent, they use *NCITED*, denoting the number of patents cited by the original patent. Only one backward-looking measure has a different name then its forward-looking counterpart: Originality. It is computed by the equation in figure 4 and is basically generality, but then looking backwards. The more original a patent is, the broader the technological roots of the underlying research. The idea behind this is: “the synthesis of divergent ideas is characteristic of research that is highly original and hence basic in that sense” (Trajtenberg, Henderson & Jaffe, 1997).

Figure 4

The formula for calculating the originality of a patent

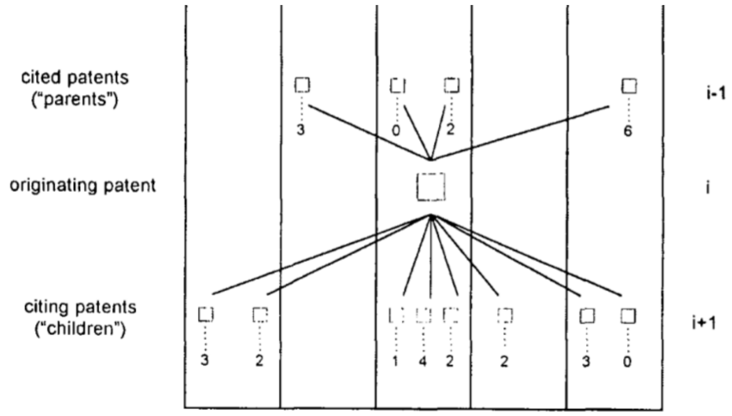
$$ORIGINAL_i = 1 - \sum_{k=1}^{N_i} \left(\frac{NCITED_{ik}}{NCITED_i}\right)^2$$

Source: Trajtenberg, Henderson & Jaffe (1997).

By using originality and generality as measures for the quality of patents, forward- and backward-looking aspects are captured. In figure 5 a graphical representation of generality and originality are provided.

Figure 5

Graphical representation of cited and citing patents, originality and generality



$$GENERAL = 1 - \left[\left(\frac{2}{8}\right)^2 + \left(\frac{3}{8}\right)^2 + \left(\frac{1}{8}\right)^2 + \left(\frac{2}{8}\right)^2\right] = \frac{23}{32}$$

$$ORIGINAL = 1 - \left[\left(\frac{1}{4}\right)^2 + \left(\frac{2}{4}\right)^2 + \left(\frac{1}{4}\right)^2\right] = \frac{5}{8}$$

Source: Trajtenberg, Henderson & Jaffe (1997).

After introducing these measures, Trajtenberg, Henderson & Jaffe use them to analyse whether university research is more basic than corporate research. They find that this is indeed the case. Along the way, they also conclude via econometric analysis that the forward- and backward-looking variables are good measures of basicness, appropriability and thus quality of a patent.

Hall, Jaffe and Trajtenberg (2001) published a paper on the dataset they had been compiling for the last decade with 30 years of patent data from US firms. They constructed measures of generality and originality based on the concepts and formulas introduced by Trajtenberg, Henderson & Jaffe (1997). In calculating these values, they made the concept more applicable to their data. They used percentage of citations received by patent I that belong to patent class j, out of n_i patent classes. The formula is depicted in figure 6. This also goes for originality, except they used backward-looking citations instead of forward-looking.

Figure 6

Formula for generality

$$Generality_i = 1 - \sum_{j}^{n_i} s_{ij}^2$$

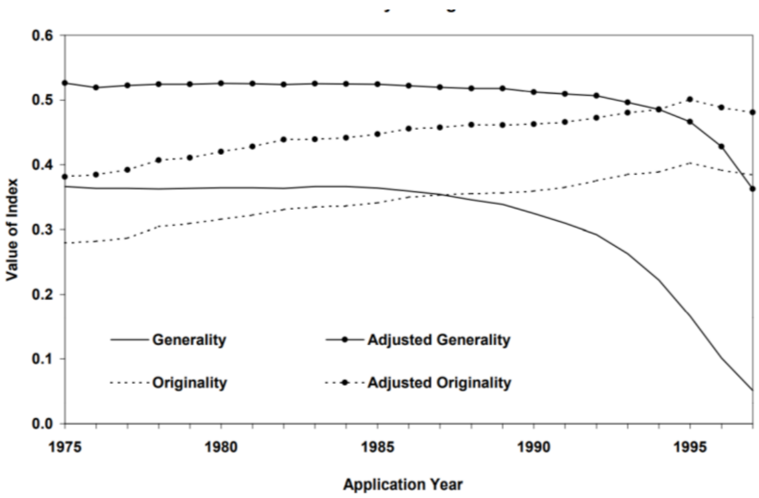
Hall, Jaffe and Trajtenberg (2001).

Hall, Jaffe and Trajtenberg (2001) make a couple of relevant statements regarding these two measures. In their dataset, the variable generality will have a high value when the original patent is cited by subsequent patents that belong to a wide range of fields. Therefore, a high score means that the patent has widespread impact in a variety of technological fields. For the originality variable in their dataset, it goes the other way around. If a patent cites previous patents that belong to a narrow set of technologies the originality score will be low. The score will be high when cited patents belong to a wide range of fields.

Figure 7 is taken from their working paper and shows the course of generality and originality.

Figure 7

Measures of generality and originality: yearly averages



Source: Hall, Jaffe and Trajtenberg (2001).

2.4 Hypotheses

Existing literature shows that patents provide a good indication of innovation, but the causal effect patents have on innovation is ambiguous. Also, patent quality positively influences firm performance. Patent quality is indicated by basicness measures such as generality and originality. The question still remains whether the number and quality of patents by incumbent firms influences the number of patents applied for by new entrants. This thesis tries to answer this question. The research question of this thesis leads to four hypotheses that together provide a complete picture of the debate of the effect patents from incumbents have on patents from entrants.

The first relationship analysed in this thesis is between the number of patents by incumbents and the number of patents by entrants. On the one hand, it is reasonable to assume that the number patents applied for by incumbent firms has a positive effect on the number of patents from entrants. Patenting ensures that more valuable information and innovations become known to the world. If more incumbents apply for patents, more information regarding innovation will be disclosed. This can inspire other individuals and firms to innovate themselves. Also, newly disclosed information could be the final piece of the puzzle allowing new firms to lay the final hand on their own innovations and breakthroughs.

On the other hand, the relationship could also be expected to be negative. If incumbents apply for more patents, these patents will cover a larger area of technological space, creating a natural patent pool. As described by Lampe and Moser (2010), patent pools discourage innovation. Also, Moser (2016) proved that when there are more patents, the risk of overlap is higher, increasing risk of litigation. This discourages other firms from entering the market, thus lowering innovation due to more patents from incumbents. This thesis tests whether this relationship is negative or positive. Therefore, the first hypothesis will be:

***H1:** the number of patents applied for by incumbents has a negative effect on the number of patents applied for by entrants.*

For the quality of patents the expectation is different. If the quality of patents is high, the positive effect of patents from entrants will kick in. Patents with higher quality, meaning higher generality and originality scores, will inspire firms to innovate further. They will also choose for patenting instead of secrecy, because a new firm cannot stay behind on other top firms. If a new firm does not patent, it appears to others as if they do not innovate.

Patent quality is measured here by generality and originality. If a patent is general, it means that it is quoted by a lot of patents in a wide range of fields in later years. Quoting a patent is the equivalent to building upon the findings in that patent. So if a patent is general, it serves as a starting point for new innovations and stimulates others to build upon the finding, thus leading to a positive relationship between quality of patents and number of patents by entrants. The same goes for originality. If a patent is original, it is built upon a wide base of earlier patents. That means that the patent combines findings from a wide range of fields. This allows other inventors to see relationships between innovations that they have not seen before, thus stimulating new innovation and patenting. Therefore, the second hypothesis will be:

H2: *the quality of patents, measured in generality and originality, applied for by incumbents has a positive effect on the number of patents applied for by entrants.*

However, hypothesis three predicts a negative relation, because when patents belonging to incumbents have higher levels of generality and originality, they are also more basic and thus crowd out the innovations of entrants. Basic patents tend to be broad, thus leading to the negative effect on innovation as seen earlier with patent pools. Also, if the firms that have the highest quality patents are incumbents, then there is little space left for new entrants to come up with and patent new innovations. Therefore, the third hypothesis will be:

H3: *the interaction terms of number of patents by incumbents and respectively generality and originality, have a more negative effect on the number of patents applied for by entrants in industries where patenting is more common.*

Lastly, the causal effects discovered via analysis will differ per technological category. Lanjouw and Schankerman (2004) displayed patent quality-adjusted measures for five different industries from 1980 to 1993. The graphs differed substantially for each industry. This gives reason to believe that the overall effects from the regression analysis actually differ per industry. Therefore, the final hypothesis will be:

H4: *There is a difference between the effects for each industry patent category.*

3. Data

3.1 Constructing the dataset

The data used for the analysis of the effect of the number and quality of patents applied for by incumbent firms on new entrants is collected from the National Bureau of Economic Research (NBER) patent data project (Hall, Jaffe and Trajtenberg, 2001). This is a well-known program that provides detailed information regarding over 3 million patents granted between 1901 and 2006. The raw data was obtained from their website.

The datasets created by NBER contain too much information divided over several datasets, so a new dataset had to be constructed for the analysis in this thesis. First, two datasets were merged. The first dataset consisted of all utility patents granted between 1901 and 2006 in multiple countries. The second dataset contained generality, originality and citation variables regarding patents in the United States. The utility patent dataset had a substantially higher amount of observations. After merging the two datasets on patent assignee number (*pdpass*) all observations from the utility dataset that did not have matching generality and originality variables were dropped. The dataset that remained had over 1.4 million observations of patents granted in the United States. Important to note is that these are not only patents applied for by US corporations, but also by universities, research institutions and the US government.

The constructed dataset is a double panel dataset based on patent holder number, which is unique for each company. The assignee dataset by NBER matches *pdpass* number to company name, but that is not relevant for this thesis (Bessen, 2009). This posed a problem. The dataset consisted mostly of duplicates. For each class there should be only one observation per year. By forcing the statistical program to drop duplicates of class and application year, the program knows that the dataset should be a panel with patent class as the panel variable and application year as the time variable. Also, all observations before 1976 were dropped, to narrow the scope of this research and allow it to be more applicable on modern day circumstances. These actions dropped a large amount of observations, leaving the dataset in a panel with 12293 observations of patents applied for in the United States between 1976 and 2006. The number of observations is the result of 30 years for each patents class, because this thesis looks at statistical relationships at the patent class level. There are 424 observed patent classes, each containing 30 observations for each year, which should bring the total number of observations to 12720. However, each class does not have 30 observations, that is simply the maximum number of observations per class. Some classes do not have observations for each year between 1976 and 2006. Therefore, the actual number of observations is slightly lower, namely 12293.

In order to analyse whether patents by incumbents influence the number of patents by entrants, a few variables had to be created. Firstly, the variable minimum year. This showed the first year in which a company applied for a patent. Then a dummy variable was introduced with the value 1 if a certain year was the first year in which a certain company applied for a patent. So if the year is 1985 and in that year a company applied for their first patent, the dummy variable takes on value 1. Then the number of entrants per year per patent class is determined by adding all the dummy variables for that year and patent class. The rest of the patents applied for in that year and in that patent class belong to incumbents. Now the dataset displays how many patents were applied for by incumbents and entrants for each year and patent class.

In conclusion, the panel dataset is sorted on patent class. In the dataset, the viewer would see all the patents belonging to patent class one, with the corresponding company number, application number, minimum year for that company, the dummy variable whether that year is the minimum year and so on. Beneath all the patents from patent class one, the same list continues for patents class two.

3.2 Variables

Due to the extensiveness of the NBER dataset, several variables were dropped that would lead to multicollinearity. The variables that remain are presented in table 1:

Table 1

List of variables used in this thesis

Variable name	**Explanation**
Incumbent firm	Firm that has already applied for a patent in an earlier year than the observed year
Entrant	Firm that applied for their first patent in the observed year
Appyear	Application year; year in which the patent was applied for
cat	Patent category; Hall, Jaffe and Trajtenberg technology category (1-6), see figure X
nclass	Patent class; US 3-digit current classification code
patent	Patent number; 7-digit patent number
pdpass	Unique patent assignee number; each company has its own *pdpass* number
general	Measure of forward-looking citations, indicates the extent to which the follow-up technical advances are spread across multiple technological fields
original	Measure of backward-looking citations, indicates how broad the technological roots are of the underlying research

nciting	Number of patents citing the original patent
ncited	Number of patents cited by the original patent
minyear	Minimum year; the first year a company (identified via unique *pdpass* number) applied for a patent
d_entry_year	Dummy variable; takes on value 1 when the year is equal to the minimum year
n_entrants	Number of entrants; the number of patents applied for in a certain year by companies for which the *pdpass* number applied for the first patent, so the dummys for *minyear* added together, that are from the same class.
N_pats_incumbents	Number of patents from incumbents; the number of patents applied for in a certain year, for the *pdpass* number was not the first patent, that are from the same class
all_pats	All patents; all patents applied for in a certain year and class
n_pats_entrants	Number of patents by entrants;
int_incumb_gen	Interaction effect variable between incumbents and generality
int_incumb_orig	Interaction effect variable between incumbents and originality
int_incumb_ncited	Interaction effect variable between incumbents and backward-looking citations
int_incumb_nciting	Interaction effect variable between incumbents and forward-looking citations

Table 2

Hall, Jaffe Trajtenberg technology categories with corresponding patent classes

Cat. Code	Category Name	Sub-Cat. Code	Sub-Category Name	Patent Classes
1	Chemical	11	Agriculture, Food, Textiles	8 19. 71. 127. 442, 504 106,118, 401, 427
		12 13	Coating Gas	48.55.95.96
		14	Organic Compounds	534,536 540. 544. 546, 548 549,552,554.556,558,560, $62,564,568.570
		15	Resins	520, 521. 522, 523, 524,525. 526,527,528.530
		19	Miscellaneous-chemical	23.34,44.102117,149,156, 159. 162 196, 201, 202 203, 204,205,208,210,216,222, 252.260,261.349.366,416 422,423,430,436,494,501, 502.510.512.516,518,585, 588
2	Computers & Communications	21	Communications	1178.333.340.342.343. 358, 367 370, 375 379, 385 455
		22	Computer Hardware & Software	341, 380. 382. 395, 701, 702. 704, 705. 706. 707 708,
		23 24	Computer Peripherals Information Storage	709. 710. 712. 713. 714 345. 347 360, 365 369, 711
3	Drugs & Medical	31	Drugs	424, 514
		32	Surgery Medical Instruments	128 600, 601, 602, 604, 606, 607
		33	Biotechnology	435,800
		39	Miscellaneous - Drug&Med	351. 433 623
4	Electrical & Electronic	41	Electrical Devices	174, 200. 327. 329, 330, 331, 332. 334, 335. 336. 337. 338. 392,
		42 43 44	Electrical Lighting Measuring Testing Nuclear X-rays	.439 313,314 315 362, 372 445 73. 324, 356, 374 250, 376 378
		45	Power Systems	60. 136, 290. 310, 318 320 322,323.361,363,388,429
		46	Semiconductor Devices	257 326 438 505
		49	Miscellancous Elec	191. 218. 219, 307. 346, 348, 377,381,386
5	Mechanical	51	Materials Processing & Handling	65, 82. 83. 125 141 142 144. 173 209, 221, 225 226, 234, 241 242. 264 271 407 408 409 414.425.451.493
		52	Metal Working	29 72 75 76 140 147 148 163. 164. 228, 266, 270, 413 419,420
		53	Motors, Engines Parts	191. 92. 123. 185. 188. 192. 251, 303 415, 417 418. 464, 474, 475,476, 477
		54	Optics	352. 353, 355, 359. 396, 399
		55	Transportation	104. 105. 114 152 180. 187 213. 238, 244, 246. 258, 280, 293, 295, 296, 298. 301, 305, 410,440
		59	Miscellancous-Mechanical	17 16, 42. 49. 51 74 81, 86. 89 100 124, 157 184 193, 194, 198.212,227,235.239,254 267.291,294.384.400,402, 406. 411, 453 454. 470, 482, 483 492, 508
6	Others	61	Agriculture, Husbandry Food	43. 47. 56 99 III. 119, 131, 426 449, 452 460
		62	Amusement Devices	273 446, 463 472. 473
		63	Apparel Textile	2, 12, 24, 26. 28 36, 38 57. 66 68 69 79 87, 112. 139, 223, 450
		64	Earth Working Wells	37 166, 171, 172 405
		65	Furniture House Fixtures	30. 70,132,182,211.256 297.312
		66	Heating	110 122. 126.165.237.373 431,432
		67	Pipes Joints	138,277, 285, 403
		68	Receptacles	53. 206. 215, 217.220.224 229 232. 383
		69	Miscellaneous- Others	11, 14. 15. 27. 33. 40. 52. 54. 59. 62. 63. 84. 101. 108. 109 116. 134. 135. 137. 150, 160, 168, 169,177.181,186,190,199, 231.236,245,248.249,269, 276,278,279,281.283,289 292. 300.368.404.412,428, 434.441.462.503

Source: Hall, Jaffe & Trajtenberg (2001).

In table 2 an overview is provided of the six patent categories Hall, Jaffe and Trajtenberg (2001) introduced. In the rightmost column the patent classes belonging to each patent category have been listed.

In the appendix A the descriptive statistics for the relevant variables are provided.

4. Methodology

This chapter will present and explain the research approach used to answer the research question and to assess whether the hypotheses can be either accepted or rejected. Several different models were used to gain an overall view of the statistical relationship between number of patents by entrants and its determinants. These determinants are number of patents by incumbents and quality measures of basicness, for example. The formulas for each regression analysis will be displayed and explained.

4.1 Linear regression analysis

This thesis presents several regression models which investigate the determinants of the number of patents applied for by first-time patentors. The main aim of regression models is to formulate the relationship between the dependent and the independent variables, based on statistical methods (Sykes, 1993). The regression models are built using the data and variables described in the previous chapter. The abbreviations used in the formulas below are listed and explained in table 1 in the previous chapter

Firstly, a linear regression with number of patents applied for by entrants as the dependent variable and the number of patents applied for by incumbents as the independent variable was performed using STATA software. The model is depicted in formulas 1 and 2.

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \varepsilon_{it}(1)$$

$$with\ t = 1{,}2 \ldots.(1902{,}2006)\ and\ \varepsilon_{it}\ being\ the\ error\ term$$

$$Patents\ by\ entrants_t = \beta_0 + \beta_1 patents\ by\ incumbents_t + \varepsilon_{it}\ (2)$$

Secondly, a linear regression model was created that controlled for different factors, such as quality measures of all patents in the database. Generality and originality are for example indicators of the quality of patents. These variables were added to the second regression in order to quantify their effect on the relationship researched in this thesis. To control adequately control for these variables, mean generality and mean originality were used. The model is depicted in formula 3 .

$$Patents\ by\ entrants_t = \beta_0 + \beta_1 patent\ by\ incumbents_t + \beta_2 general + \beta_3 originality + \varepsilon_{it}\ (3)$$

$$with\ t = 1{,}2 \ldots.(1902{,}2006)$$

4.2 Linear regression analysis with interaction effects

The third model used in this thesis is a linear regression model with interaction effects. The four interaction effects used in this thesis are between the number of patents by incumbents and generality, originality, number of forward-citations and number of backward-citations. Formally, an interaction occurs when an independent variable has a different effect on the outcome depending on the values of another independent variable. When using interaction effects, the relationship between number of patents from entrants depends on the mean generality of patents from incumbents versus mean generality of all patents. Therefore, this method allows the estimation of the effect of quality of incumbent patents. If the interaction terms are significant, it will indicate that effect of number of patents applied for by incumbents on number of patents applied for by entrants is different at different values of generality and originality of the incumbent patents. The interaction effect model is depicted in formula 4.

$$Patents\ by\ entrants_t = \beta_0 + \beta_1 patent\ by\ incumbents_t + \beta_2 general_t + \beta_3 originality_t + \beta_4 nciting_t + \beta_5 ncited_t + \beta_6 interaction\ incumbent\ gen_t + \beta_7 interaction\ incumbent\ orig_t + \beta_8 interaction\ incumbent\ ncited_t + \beta_9 interaction\ incumbent\ nciting_t + \varepsilon_{it} \quad (4)$$

4.3 Fixed effects regression analysis

The last model used is a fixed effects analysis model was used. This type of regression allows controlling for time-invariant characteristics when the potential control variables are not available or cannot be identified. In this thesis, the relationship between the amount of patents applied for by first-time patenting entities and the amount and quality of patents applied for by experienced patenting entities might be influenced by time-invariant characteristics differing between companies. These time-invariant factors could be patent law, wage levels per industry, foreign direct investments in the United States and political system (Chatelain & Ralf, 2021). These variables are relatively fixed over time, yet they could influence the amount and quality of patents applied for by first-time patenting companies one way or another. Therefore, fixed effects regression is used to reduce the omitted variable biases of not accounting for these time-invariant variables. The added relevance is that the fixed-effects model transforms the variables by taking time averages. Variables such as patents by entrants, patents by incumbents and all the interaction terms will be averaged out from 1901 until 2006. Thus, the model will control for time-invariant effects, which is necessary because the data is collected in the period of more than a century. Due to the design of the dataset, a panel with patent class as the panel variable, patent class fixed effects are used to control for time-invariant differences.

Using fixed effects is essential when analysing patent data. Hall, Jaffe and Trajtenberg (2001) use this method as a way to rescale the citation data. Due to time-invariant characteristics, it is difficult to compare a 1970 patent with three citations with a 2004 patent with 78 citations. A lot has changed between 1970 and 2004, too much to control for in a linear regression.

The fixed effects model is estimated in formulas 5 and 6.

$$Y_{it} = \alpha_i + \beta_1 X_{it} + \cdots + \beta_k X_{k,it} + \varepsilon_{it} (5)$$

$$with\ \alpha_i\ being\ a\ constant\ that\ differs\ per\ patent$$

$$Patents\ by\ entrants_t = \alpha_i + \beta_1 patent\ by\ incumbents_t + \beta_2 general_t + \beta_3 nciting_t + \beta_4 ncited_t + \beta_5 originality_t + \beta_6\ interaction\ incumbent\ gen_t + \beta_7 interaction\ incumbent\ orig_t + \beta_8 interaction\ incumbent\ ncited_t + \beta_9 interaction\ incumbent\ nciting_t + \varepsilon_{it}\ \text{(6)}$$

4.4 Analysis whether results differ per industry

As mentioned in the theoretical framework, Lanjouw and Schankerman (2004) displayed patent quality-adjusted measures for five different industries from 1980 to 1993 and the graphs differed substantially for each industry. This gives reason to believe that the overall effects from the regression analysis actually differ per industry. To test this, the model depicted in formula 4 was used six times, each on a dataset that only held observations belonging to one category of industry.

5. Results

This chapter will present and interpret the most relevant results from the regression analyses described in the previous chapter. The following indicators of significance are used in the tables presented in this chapter and the appendix. One star indicates significance at the 5% level, two stars indicates significance at the 1% level and three starts indicates the significance at the 0.1% level.

5.1 Linear regression model results

The results of the first two models mentioned in the Methodology chapter provide insights on the first two hypotheses. These hypotheses state that the number of patents from incumbents has a negative effect on the number of patents from entrants, while the mean quality of patents has a positive effect. The results are displayed in table 3.

Table 3

Results from linear regression models 1 and 2

	(1) n_pats_entrants	(2) n_pats_entrants
n_pats_incumbents	0.149***	0.144***
	(0.0140)	(0.0159)
Generality		1.121
		(0.591)
Originality		0.294
		(0.404)
Constant	12.75***	13.38***
	(1.047)	(1.423)
Observations	12268	6410

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The first column shows the outcomes of the model using formula 2 and the second column shows the same for formula 3. The coefficient belonging to the number of patents by incumbents (*n_pats_incumbents*) is positive and significantly differs from zero at a 0.1% level. However, the sign does become smaller in model 2, indicating that the effect of the number of patents by incumbents becomes smaller when adding control variables. The coefficient can be interpretated as follows: an increase in the number of patents from incumbents by 1 increases the number of patents from entrants by 0.149 and in model 2 by 0.144. These results show that the relationship is positive between the independent and dependant variable, thus meaning that the first hypothesis must be rejected.

Model 2 shows a positive and non-significant coefficient for mean generality and mean originality. This indicates that there is not enough statistical evidence to assume that generality and originality have a significant effect on the number of patents by entrants. If generality for example were to be significant, the interpretation would be that if the mean generality score of all patents in the dataset increases by 1, than 1.121 more patents are applied for by entrants. Still the sign of both coefficients is positive. This indicates, rather than proves, that increased generality and originality of patents positively influences new firms to innovate and patent their innovations. However, due to non-significance it still means that hypothesis 2, that the quality of patents, measured in generality and originality, applied for by incumbents has a positive effect on the number of patents applied for by entrants, must be rejected. Perhaps when a larger dataset is used, significance can be achieved for generality and originality. This thesis only shows an indication of their influence instead of providing evidence.

5.2 Linear regression model with interaction variables results

In table 4 the results for the third regression model from formula 4 are displayed along with the earlier models incorporated to allow easy comparison between models. This table can be used to draw conclusions regarding hypothesis 3, whether the interaction terms of number of patents by incumbents and respectively generality and originality, have a *more negative* effect on the number of patents applied for by entrants in industries in which patenting is more common. The coefficients of the interaction terms between number of patents from incumbents and generality (*int_incumb_gen*) and forward citations (*int_incumb_nciting*) are the only significant interaction terms, the other two are non-significant. The sign is positive for the interaction term with generality, indicating that when incumbents patents are more general, 0.0330 more patents will be applied for by entrants. This is not in line with the third hypothesis. The effect is small, but still significant. The sign is negative for the interaction term with forward citations as for the sign of forward citations itself. This is in line with the third hypothesis. It also means that it is safe to assume that the researched relationship becomes more negative in industries where patenting is more common. More patents means that the effect of patents becomes larger.

Therefore, according to model 3 the third hypothesis must be partially rejected. However, one must keep in mind that this outcome is not very reliable. That is because only two of the four interaction terms are significant. Two of the four showed a negative sign. Even though these interaction terms are non-significant, they do indicate that hypothesis 3 cannot blindly be fully rejected. The only evidence for the third hypothesis lays within the significance and negative sign of the interaction term between number of patents from entrants and number of forward citations.

Table 4

Results from linear regression models 1,2 and 3

	(1) n_pats_entrants	(2) n_pats_entrants	(3) n_pats_entrants
n_pats_incumbents	0.149***	0.144***	0.142***
	(0.0140)	(0.0159)	(0.0199)
Generality		1.121	-1.771
		(0.591)	(1.281)
Originality		0.294	1.688
		(0.404)	(1.211)
Total forward cites to the patent			-0.0423**
			(0.0158)
Total backward cites from the patent			0.156***
			(0.0281)
int_incumb_gen			0.0330*
			(0.0131)
int_incumb_orig			-0.0193
			(0.0162)
int_incumb_ncited			0.000144
			(0.000194)
int_incumb_nciting			-0.000239**
			(0.0000736)
Constant	12.75***	13.38***	13.13***
	(1.047)	(1.423)	(1.687)
Observations	12268	6410	6410

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

5.3 Fixed effects regression results

The first three hypotheses are also analysed using the model from formula 6, an individual fixed effects regression, based on patent class fixed effects. With this tool, the hypotheses will be assessed from an OLS and fixed effects viewpoint. This provides a more complete answer to the research question.

In table 5, the output of the fixed effects regression is displayed. The coefficient for number of patents by incumbents (*n_pats_incumbents*) still is significant and positive and roughly the same size. This means that the fixed effects model also rejects hypothesis 1. The coefficients for mean generality and mean originality are both not significant. This still results in a rejected second hypothesis. However, backward and forward citations to patents were included. Backward citations has a positive and significant effect. That means that if patents contain on average one more citations to previous patents, 0.152 more patents will be applied for by entrants. Forward citations has a negative and significant effect. This means that if all patents are cited on average one more time in later patents, indicating that the patent provided seminal technological breakthroughs, 0.0456 less patents will be applied for by entrants. Therefore, the fixed effects model only partially rejects hypothesis 2, because it shows some evidence via forward and backward citations.

Conclusions about the third hypothesis, that the four interaction effects have a more negative influence on the number of patents from entrants in industries in which patenting is more common, are the same when using a fixed effects model. The interaction effect between incumbent and generality is still positive, significant and roughly the same size as in OLS regression. The same goes for the interaction effect with forward citations. The first interaction effect still does not support the third hypotheses, because it indicates a positive relationship between incumbent patent quality and number of patents from entrants. The interaction effect between incumbent and number of forward citations (*nciting*) being negative and significant does support the second hypothesis. Therefore, hypothesis still three cannot be fully rejected, only partially. The other two interaction effect variables remain non-significant, thus no conclusions can be drawn from their values.

Table 5

Results from fixed effects regression

	(1) n_pats_entrants
n_pats_incumbents	0.142***
	(0.0202)
Generality	-1.732
	(1.275)
Originality	1.566
	(1.201)
Total forward cites to the patent	-0.0456**
	(0.0156)
Total backward cites from the patent	0.152***
	(0.0278)
int_incumb_gen	0.0306*
	(0.0127)
int_incumb_orig	-0.0192
	(0.0159)
int_incumb_ncited	0.000153
	(0.000188)
int_incumb_nciting	-0.000228**

	(0.0000709)
Constant	14.24***
	(1.888)
Observations	6410

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

5.4 Test whether effects differ per industry

Hypothesis 4 states that there is a difference between the effects described in this chapter for each category of industry. These category are HJT technology categories, based on Hall, Jaffe and Trajtenberg (2001). The six categories are listed in table 2.

In order to test this hypothesis, the regression in formula 4 was performed six times, each time with an edited dataset so it only contained observations where the category was the same. The results are displayed below in table 6. The regression tables can be found in appendix C. In this chapter only a summary table is presented to illustrate the most relevant results regarding hypothesis 4.

Table 6

Overview of all variables in a regression with one category in the dataset

	Category 1	**Category 2**	**Category 3**	**Category 4**	**Category 5**	**Category 6**
n_pats_incumbents	0.140***	0.163***	0.161***	0.0797***	0.187***	0.261***
Generality	-1.219	4.961	0.841	-7.017**	-2.915***	-0.668
Originality	-0.377	-0.779	10.85*	2.680	0.688	0.296
int_incumb_gen	0.0103	0.00364	0.00643	0.0727***	0.0844***	0.00525
int_incumb_orig	0.00399	0.00156	-0.0736**	-0.0104	-0.00261	0.000656
int_incumb_ncited	-0.000245*	0.000601*	0.000333***	-0.000294**	-0.000210	-0.000276
int_incumb_nciting	0.000922*	0.0000676	0.000790***	-0.000175*	-0.0000321	-0.000729**
Observations	1080	880	224	878	1616	1732

It is clear that there are large differences in sign, size and significance per variable in each category. For example, generality has a positive sign in two categories and negative in the other four. Significance also differs widely, as does size. Only the variables number of patents from incumbents (*n_pats_incumbents*) and the interaction effect between incumbents and generality (*int_incumb_gen*) keep the same sign and roughly the same size in each category. *N_pats_incumbents* even remains highly significant in all categories.

However, most variables are non-significant, making it difficult to draw conclusions. Still, this table indicates, rather than proves, that the effect of these variables on the number of patents from entrants differs when the patent belongs in a different HJT technological category. The effect is only mostly the same for number of patents from incumbents and the incumbent generality interaction effect. Therefore, hypothesis four is largely accepted.

6. Conclusion and Discussion

This bachelor thesis researches how the number and quality of patents from incumbents effects the number of new firms applying for their first patent. It uses a panel dataset from NBER with 12293 observations regarding patents applied for in the United States between 1976 and 2006. One of the main findings is that more incumbents applying for patents stimulates companies to apply for their first patent. In this chapter, four key conclusions are drawn from this research. Also, their implications for economic theory and patent policy are explained. Finally, shortcomings and suggestions for further research are described.

6.1 Conclusions hypotheses 1 and 2

The results of the first regression model, displayed in table 3, indicate that the number of patents from incumbents has a positive effect on the number of patents applied for by entrants. Therefore, the first hypothesis is rejected. This means that in a year in which more incumbent firms apply for a patent, it is more likely that firms will apply for their first patent. This is the most convincing conclusion in this thesis. The reason is that the variable that indicates the number of patents applied for by incumbents (*n_pats_incumbents*) remains positive and significant in every regression model. Table 6 from hypothesis 4 even shows that this variable remains positive and significant in all six patent categories.

This finding implies that the number patents applied for by incumbents has a larger positive effect than negative. The negative effect would be that more patents cover a larger area of technological space and create natural patent pools, thus discouraging innovation as stated by Lampe and Moser (2010). Apparently, disclosing more innovations inspires other to patent their own innovations.

This finding also has implications for government policy. As mentioned in the introduction and theoretical framework, governments use patent policy as a way to stimulate innovation. It has been proved that the relationship between patenting and innovation is ambiguous. However, the conclusion for hypothesis 1 suggests that by stimulating companies to patent, more companies will follow. This does not prove that innovation is stimulated. It could be the case that companies had the innovation in the pipeline and chose for secrecy. Now that competitors chose patenting, they follow. Still, if a firm chooses to patent, this will have positive spill over effects. More patents means that more innovations become known to the public. This will inspire other firms and individuals to innovate for themselves.

The second hypothesis is rejected. Models 2, 3 and 4 show a positive and non-significant coefficient for mean generality and mean originality. This indicates that there is not enough statistical evidence to assume that generality and originality have a significant effect on the number of patents by entrants. However, backward and forward citations of a patent are highly significant in the fixed effects model. This implies that generality and originality of all patents does not influence the number of patents from entrants. However, backward and forward citations, which are also quality indicators, do influence the independent variable. This is a key finding and has several implications.

Using the fixed effects model, backward citations of patents in general have a positive impact. This implies that when patents on average are based on a lot of earlier innovations, so have a high originality score, they have a positive influence on new entrants. The more original a patent is, the broader the technological roots of the underlying research. According to Halle, Jaffe and Trajtenberg (2001), a high originality score indicates basicness. This implies that the more basic a patent is measured with backward-looking indicators, the more inspiring it does. Innovative and broadly supported patents inspire companies and individuals to patent themselves, thus a positive coefficient.

The fixed effects model also shows that forward citations have a negative impact. This indicates that when the entire sample of patents is cited on average more in later patents, that this will negatively influence the number of patents applied for by entrants. The more forward citations a patent has, the higher the generality score. Therefore, generality indirectly negatively influences the independent variable in this research. An explanation is that the more forward citations a patent later collects, the more crowding out the patent does. Too many innovators are bound by the findings in the original patent. In conclusion, hypothesis two must be partially rejected.

6.2 Conclusion hypothesis 3

Hypothesis 3 predicted that the interaction terms between incumbents and quality indicators have a more negative effect on patents from entrants in industries in which patenting is more common. This hypothesis is partially rejected. This conclusion is based on the patent class fixed effects model, because it controls for time-invariant variables differing per patent class. The notion that when patents belonging to incumbents have higher levels of generality and originality, they are more basic and thus crowd out the innovations of entrants, is partially true. The interaction effect between incumbent and amount of forward citations of a patent is negative, providing proof of this notion. The interpretation is as follows: *Int_incumb_nciting* being negative means that less patents will be applied for by entrants if the patents applied for in that year by incumbents receive more forward citations in later years. If more patents are applied for in an industry, this negative effect will influence more entrants. Therefore, the negative effect will have more effect in industries with more patenting. This effect is very small. However, the results contradict each other. The interaction effect between incumbent and generality is positive. As generality is based on forward citations, this creates a contradiction. An explanation could be that more patents had corresponding forward citation data than generality data coupled to it. Still the results implicate that quality of patents from incumbents influences how many entrants apply for their first patent. The question how this influences entrants cannot be fully answered. Therefore, this third conclusion is less strong than the previous two conclusions.

6.3 Conclusion hypothesis 4

Hypothesis 4 states that there are differences in the effects of number and quality of patents from incumbents on the number of patents from entrants per industry. The results suggest that this is the case. The implication of this finding is that patent policy must differ per industry. Industries will most likely react differently to changes in patent policy and the consequences will be heterogenous. Existing literature on this topic already shows that different industries have a larger bias towards secrecy of new innovations, while others prefer patenting (Cohen, Nelson and Wash, 2000). This thesis suggests in addition to this finding that effects of quantity and quality measures of patents from incumbents on patents from entrants differ per industry.

6.4 Conclusion to the research question

The research question of this thesis is: "*How do quality and number of patents from incumbent firms impact the amount of firms applying for their first patent?"*. Four hypotheses were created and quantitatively tested using several regression models. This research was motivated by the paper from Trajtenberg, Henderson and Jaffe (1997) in which they introduce the generality and originality measures of basicness to assess the quality of patents. This thesis connected these concepts to the causal effect patents from incumbent firms have on patents from entrants and whether the former encourages the latter. This research has significant policy implications and adds to scientific research regarding patenting. The key findings are that the number of patents from entrants positively influences the number of patents by entrants in a given year. Also, forward citations of patents in general have a negative impact, while backward citations have a positive effect. These are quality measures for all patents. Generality and originality do not have direct significant effects on the number of patents from entrants. When looking into quality measures describing patents from incumbents, forward citations also has a negative effect. However, this effect is small. Generality of patents from incumbents has a small positive effect. Interestingly, generality becomes significant when interacting with the number of entrants, while lacking significance in all other models. Lastly, the outcomes of the thesis suggest that all the effects discussed differ per industry.

6.5 Shortcomings and suggestions for future research

This thesis has several shortcomings. Firstly, the quality measures for patents used in this thesis are based on citations. The issue is that there can be a bias in citations. Hall, Jaffe and Trajtenberg (2001) explain in their working paper discussing the construction of their dataset that the measures used in this thesis tend to be positively correlated with the number of citations a patent receives in general. Patents with breakthrough innovations will tend to have more citations and thus have higher generality scores. Likewise, patents that use a lot of citations would display on average higher originality. Also, when a patent receives more citations, there is a tendency that the patent covers more patent classes. This tendency can be real, but still it forms a bias in the dataset.

Secondly, one of the key findings in this thesis is that the number of patents from incumbents positively influences the number of patents by entrants. So it stimulates entrants to apply for their first patent, but it remains unsure whether number of patents by incumbents also stimulates innovation. An innovation could have already been created and the company could have chosen for secrecy, until incumbents applied for a lot of patents which stimulated them to switch from secrecy to patenting. Either way, more innovations are patented, which means more innovations become known to the public. This inspires individuals and companies to start innovating themselves. The causal effect may not have been found, but the inspiration is definitely there.

Thirdly, there are issues with using patents as a proxy for innovation. This limitation relates to the second limitation and makes it broader. This thesis itself is not bothered by that fact, but it does when commenting on the applications of this research. Certain relationships have been found in patent data. However, the question remains how applicable this is in the discussion of innovation stimulation. For example, Moser (2016) and Acs and Audretsch (1988) suggest that patents do not stimulate innovation. Thus making this thesis not applicable to the innovation stimulation discussion. Still, patents remain a good measure for innovation.

Lastly, generality and forward citations should have the same sign, because they are based on each other. As shown in the fixed effects regression, generality is negative, forward citations is negative and the interaction effect between incumbent and generality is positive. As generality is based on forward citations, this creates a contradiction. An explanation could be that the formula to compute the generality measure crosses away negative signs in the effect of forward citations or vice versa. Still, this limitation questions the validity of this research.

For further research hypothesis four can be tested more accurately. This goes beyond the scope of this thesis, but it remains an interesting topic to analyse the differences in the effects of patenting between industries. Secondly, further research can use time lags to assess the impact of forward citations. As mentioned earlier, the effect forward citations has on number of patents from entrants differs when looking at the entire sample of patents or just incumbents. This makes the effect rather ambiguous. By using time lags in future research, the effect of forward citations may be better described. Forward citations are gathered in the years after the patent is applied for. Therefore, time lags may provide more insight on the effects forward citations have on other patent-related variables. Lastly, further research that may be interesting from a corporate standpoint is focussing more on appropriability, rather than basicness and measuring whether you can make money from an innovation. This could provide insights for companies on how to allocate resources towards research and development.

List of References

Acs, Z. J., & Audretsch, D. B. (1988). Innovation in large and small firms: an empirical analysis. *The American economic review, 78*(4), 678-690.

Acs, Z.J., Anselin, L. & Varga, A. (2002). Patents and innovation counts as measures of regional production of new knowledge. *Research Policy, 31*(7), 1069-1085.

Arrow, K.J. (1962). Economic Welfare and the Allocation of Resources for Inventions. in R. Nelson (ed.) The Rate and Direction of Inventive Activity, Princeton University Press.

Bessen, J. (2009). NBER PDP project user documentation: Matching Patent Data to Compustat Firms. Patent Data Project - DOWNLOADS (google.com)

Chatelain, J. B., & Ralf, K. (2021). Inference on time-invariant variables using panel data: A pretest estimator. *Economic Modelling, 97*, 157-166.

Cohen, W.M., Nelson, R.R. & Walsh, J.P. (2000). Protecting their intellectual assets: appropriability conditions and why US manufacturing firms patent (or not). (National Bureau of Economic Research Working Paper No. 7552).

Hall, B., Jaffe, A. & Trajtenberg, M. (2001). The NBER patent citations data file: lessons, insights and methodological tools. (National Bureau of Economic Research Working Paper No. 8498)

Katila, R. (2000). Using patent data to measure innovation performance. *International Journal of Business Performance Management, 2*(1-3), 180-193.

Kuznets, S. (1962). Quantitative aspects of the economic growth of nations: VII. the share and structures of consumption. *Economic Development and Cultural Change, 10*(2, Part 2), 1-92.

Lamoreaux, N.R. & Sokoloff, K.L. (1999). Inventive Activity and the Market for Technology in the United States. (National Bureau of Economic Research Working Paper No. 7107).

Lampe, R. & Moser, P. (2010). Do patent pools encourage innovation? evidence from the nineteenth-century sewing machine industry. *The Journal of Economic History, 70*(4), 898-920.

Lanjouw, J.O. & Schankerman, M. (2004). Patent Quality and Research Productivity: Measuring Innovation with Multiple Indicators. *The Economic Journal, 114*(495), 41-465.

Moser, P., & Voena, A . (2012). Compulsory Licensing: Evidence from the Trading with the Enemy Act. *American Economic Review, 102*(1),396-427.

Moser, P. (2013). Patents and Innovation: Evidence from Economic History. *Journal of Economic Perspectives*, 27(1), 23-44.

Moser, P. (2016). Patents and innovation in economic history. *Annual Review of Economics, 8*(1), 241-258.

Pakes, A. & Griliches, Z. (1980). Patents and R&D at the firm level: a first look. (National Bureau of Economic Research Working Paper No. 561).

Rao, S., Ahmad, A., Horsman, W., & Kaptein-Russell, P. (2001). The importance of innovation for productivity*. International Productivity Monitor, 2*(11), 1-59.

Roger, M. (1998). *The definition and measurement of innovation.* (Melbourne Institute of Applied Economic and Social Research Working Paper No. 10/98).

Sterzi, V. (2013). Patent quality and ownership: An analysis of UK faculty patenting*. Research Policy, 42*(2), 564-576.

Sykes, A.O. (1993). An introduction to regression analysis. (Coase-Sandor Institute for Law & Economics Working Paper No. 20).

Trajtenberg, M., Henderson, R., & Jaffe, A. (1997). University versus corporate patents: A window on the basicness of invention. *Economics of Innovation and new technology, 5*(1), 19-50.

Tseng, C.Y. & Wu, L.Y. (2007). Innovation quality in the automobile industry: measurement indicators and performance implications. *International Journal of Technology Management, 37*(1-2), 162-177.

Appendix

Appendix A: Descriptive Statistics

Table 7

Descriptive statistics of each variable used in the dataset in this thesis

Variable	Obs	Mean	Std. Dev.	Min	Max
appyear	12293	1990.572	8.694	1976	2006
cat	12268	4.017	1.871	1	6
nclass	12268	309.91	186.689	1	800
patent	12293	5359868.7	949777.81	3980120	7155733
pdpass	12293	10918490	936420.03	10030609	23152592
general	8708	.581	.324	0	1
ncited	12293	9.099	17.35	0	536
nciting	12293	8.258	14.81	0	367
orig	9542	.548	.342	0	1
minyear	12293	1983.796	10.346	1903	2006
d entry year	12293	.349	.477	0	1
n entrants	12293	26.099	34.056	0	369
n pats incumbents	12293	86.157	167.968	0	2838
all pats	12293	112.256	195.065	1	3124
n pats entrants	12293	26.099	34.056	0	369
int incumb gen	8708	47.066	100.182	0	2551.5
int incumb orig	9542	49.874	114.703	0	2677.5
int incumb ncited	12293	834.652	3653.448	0	159668
int incumb nciting	12293	928.452	6288.826	0	561968

Appendix B: List of STATA commands used in the analysis

Constructing the dataset:

- duplicates drop nclass appyear, force
- bysort company nclass: egen minyear=min(appyear)
- gen d_entry_year=year==minyear
- bysort appyear nclass: egen n_entrants=sum(d_entry_year) (*to count the number of entrants per year*)
- bysort nclass appyear: egen n_pats_incumbents=sum(!d_entry_year)
- bysort nclass appyear: egen all_pats=sum(1)
- gen n_pats_entrants=all_pats-n_pats_incumbents

Hypothesis 1:

- essto: xtreg n_pats_entrants n_pats_incumbents, robust
- esttab using regression1.rtf, label se

Hypothesis 2:

- essto: xtreg n_pats_entrants n_pats_incumbents general orig , robust
- esttab using regression2.rtf, label se

Hypothesis 3:

- essto: xtreg n_pats_entrants n_pats_incumbents general orig ncited nciting int_incumb_gen int_incumb_orig int_incumb_ncited int_incumb_nciting, robust
- esttab using regression3.rtf, label se

Hypothesis 4:

- drop if cat==1, drop if cat==2, drop if cat==3, drop if cat==4, drop if cat==5 drop if cat==6 (*perform the regression six times with each time all categories dropped except one)*
- essto: xtreg n_pats_entrants n_pats_incumbents general orig ncited nciting int_incumb_gen int_incumb_orig int_incumb_ncited int_incumb_nciting, robust
- esttab using regression4.rtf, label se

Appendix C: Tables used for hypothesis 4

	Category 1	**Category 2**	**Category 3**	**Category 4**	**Category 5**	**Category 6**
n_pats_incumbents	0.140^{***}	0.163^{***}	0.161^{***}	0.0797^{***}	0.187^{***}	0.261^{***}
Generality	-1.219	4.961	0.841	-7.017^{**}	-2.915^{***}	-0.668
Originality	-0.377	-0.779	10.85^{*}	2.680	0.688	0.296
int_incumb_gen	0.0103	0.00364	0.00643	0.0727^{***}	0.0844^{***}	0.00525
int_incumb_orig	0.00399	0.00156	-0.0736^{**}	-0.0104	-0.00261	0.000656
int_incumb_ncited	-0.000245^{*}	0.000601^{*}	0.000333^{***}	-0.000294^{**}	-0.000210	-0.000276
int_incumb_nciting	0.000922^{*}	0.0000676	0.000790^{***}	-0.000175^{*}	-0.0000321	-0.000729^{**}
Observations	1080	880	224	878	1616	1732

Table 8

Interaction variables regression with category 1

	(1) n_pats_entrants
n_pats_incumbents	0.140^{***} (0.0377)
Generality	-1.219 (1.175)
Originality	-0.377 (1.023)
int_incumb_gen	0.0103 (0.0152)
int_incumb_orig	0.00399 (0.0148)
int_incumb_ncited	-0.000245^{*}

	(0.000120)
int_incumb_nciting	0.000922*
	(0.000379)
Constant	10.09***
	(2.259)
Observations	1080

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 9

Interaction variables regression with category 2

	(1) n_pats_entrants
n_pats_incumbents	0.163***
	(0.0358)
Generality	4.961
	(5.691)
Originality	-0.779
	(2.924)
int_incumb_gen	0.00364
	(0.0359)
int_incumb_orig	0.00156
	(0.0235)
int_incumb_ncited	0.000601*
	(0.000263)
int_incumb_nciting	0.0000676
	(0.000243)
Constant	7.393
	(4.289)
Observations	880

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 10

Interaction variables regression with category 3

	(1) n_pats_entrants
n_pats_incumbents	0.161***
	(0.0246)
Generality	0.841
	(4.090)
Originality	10.85*
	(5.077)
int_incumb_gen	0.00643
	(0.00553)
int_incumb_orig	-0.0736**
	(0.0276)
int_incumb_ncited	0.000333***
	(0.0000810)

int_incumb_nciting	0.000790***
	(0.000160)
Constant	18.74***
	(3.924)
Observations	224

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 11

Interaction variables regression with category 4

	(1) n_pats_entrants
n_pats_incumbents	0.0797***
	(0.00946)
Generality	-7.017**
	(2.584)
Originality	2.680
	(1.701)
int_incumb_gen	0.0727***
	(0.0213)
int_incumb_orig	-0.0104
	(0.00968)
int_incumb_ncited	-0.000294**
	(0.000110)
int_incumb_nciting	-0.000175*
	(0.0000746)
Constant	20.80***
	(2.976)
Observations	878

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 12

Interaction variables regression with category 5

	(1) n_pats_entrants
n_pats_incumbents	0.187***
	(0.0202)
Generality	-2.915***
	(0.809)
Originality	0.688
	(0.559)
int_incumb_gen	0.0844***
	(0.0156)
int_incumb_orig	-0.00261
	(0.0143)
int_incumb_ncited	-0.000210

	(0.000225)
int_incumb_nciting	-0.0000321
	(0.000343)
Constant	11.29***
	(1.270)
Observations	1616

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 13

Interaction variables regression with category 6

	(1)
	n_pats_entrants
n_pats_incumbents	0.261***
	(0.0343)
Generality	-0.668
	(1.295)
Originality	0.296
	(0.578)
int_incumb_gen	0.00525
	(0.0304)
int_incumb_orig	0.000656
	(0.00566)
int_incumb_ncited	-0.000276
	(0.000249)
int_incumb_nciting	-0.000729**
	(0.000253)
Constant	12.20***
	(1.605)
Observations	1732

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix D: Robustness check with negative binomial regression

To assess the robustness of the results, it is important to check whether the data is robust as well. Tabulating the number of patents by entrants by patent class revealed that the conditional variance per patent class is larger than the conditional mean. This indicates over-dispersed count data. Negative binomial regression is a suitable method to assess whether this phenomenon leads to non-robustness of the results. Therefore, a negative binomial regression was used to evaluate the robustness of the results discussed earlier.

The coefficients in the negative binomial regression show the difference in expected count of number of patents by entrants per class compared to the reference group. This information is not very relevant for this thesis. More relevant is predicting the margins at each patent class. A part of the results of the margin estimation is displayed in table 14. Including all 424 classes would lead to an unorganized table, which is why most classes have been left out. The table that remains still gives a good indication of the outcomes. The coefficients can be interpreted as follows: a coefficient of 3.012 for patent class 2 means that the predicted number of patents by entrants in class 2 is 3.012. The full table shows that most coefficients are significant. This significance indicates that the negative binomial model serves as a testimony to the robustness of the results. If the conditional variance in a patent class were too large, then the coefficients would not have been significant. This would mean that no conclusions could be made about the relationship between patents by incumbents and patents by entrants for a specific patent class without a significant coefficient. A small amount of class coefficients are non-significant, which places a check on the robustness of the dataset and the results coming from them.

The p-value of the entire negative binomial model is 0.000. This indicates that the model is highly significant and that the results are robust according to a negative binomial regression analysis. In conclusion, this robustness check proves that the results are robust to negative binomial regression and that with this dataset applicable conclusions can be made regarding the relationship between number of patents by incumbents and number of patents by entrants.

Table 14

Part of the results of the margin estimation at each patent class in the negative binomial analysis

	(1) n_pats_entrants
n_pats_entrants	
n_pats_incumbents	0.00278***
	(0.0000528)
US 3-digit current classification (CCL)=1	0
	(.)
US 3-digit current classification (CCL)=2	3.012***
	(0.381)
US 3-digit current classification (CCL)=4	2.485***
	(0.381)
US 3-digit current classification (CCL)=5	2.744***
	(0.381)
US 3-digit current classification (CCL)=7	0.877*
	(0.391)
US 3-digit current classification (CCL)=8	2.140***
	(0.382)
US 3-digit current classification (CCL)=12	0.238
	(0.401)
US 3-digit current classification (CCL)=14	1.176**
	(0.387)
US 3-digit current classification (CCL)=15	2.950***
	(0.380)
US 3-digit current classification (CCL)=16	2.735***
	(0.381)
US 3-digit current classification (CCL)=19	0.903*
	(0.392)
Constant	0.511
	(0.370)
lnalpha	-1.512***
	(0.0179)
Observations	12268