Examining the relation between patent value and patent claims

Jansen, W.

Award date:
2009

Examining the relation between patent value and patent claims

Wouter Jansen

February 19, 2009

Master Thesis Report by Wouter Jansen

Supervisors

Eindhoven University of Technology:

Dr. Önder Nomaler

Dr. Alessandro Nuvolari

Philips, Intellectual Property and Standards:

Piet van Zanten

Gerard van der Ligt

Examining the relation between patent value and patent claims

February 19, 2009

Final report of a Master Thesis project at the Eindhoven University of Technology, conducted at the Philips department of Intellectual Property & Standards.

Student

Wouter Jansen

0526938

Supervisors

On behalf of the Eindhoven University of Technology, School of Innovation Sciences:

Dr. Önder Nomaler

Dr. Alessandro Nuvolari

On behalf of Philips Intellectual Property and Standards, department of Business Intelligence:

Piet van Zanten

Gerard van der Ligt

1 Summary

Patents are legal documents providing their owner with exclusive rights to commercially exploit the invention they describe. Because patents are a unique, well-organized source of data on technological development over a long time frame, researchers use them to measure e.g. innovativeness or technology trajectories. Also in recent years, businesses shifted from using patents only as a defensive mechanism protecting their own products towards active exploitation of IP rights, for instance by proactively looking for licensees or by influencing standardization processes.

1.1 The research question and research set-up

Determining the value of a patent is among the key questions in the patent world. Currently, accurate value estimations require expensive and time-consuming qualitative analyses by patent attorneys (experts in both legal and technical matters). Therefore, quantitative 'patent indicators' were developed, capable of ranking patent portfolios. These patent indicators use patent metadata from databases, like for instance citations or renewal data, to conduct statistical patent valuation. Most of the established indicators fail to take into account content-related metadata and suffer from timeliness (the fact that they become available at a rather late stage in the patent's life). Therefore, Reitzig (2004a) proposed the construction of a full-text patent indicator that can predict patent value at an early stage of the patent's life. This thesis project examined if such an indicator could be built and therefore had as its main research question:

How can full-text patent data be used to predict the value of a patent?

This question was answered by a qualitative exploration of the full-text characteristics, after which a quantitative analysis was conducted to empirically validate the relation of these characteristics with patent value. Both steps will be described hereafter.

1.2 Claim Characteristics

The qualitative exploration required a research environment that would give access to expert knowledge on patents and the use of quality databases for empirical validation. The Intellectual Property & Standards (IP&S) department of Philips provided such a platform. This gave me the opportunity to interview several patent attorneys (engineers schooled in patent law). They suggested narrowing down 'full-text' data to the analysis of the claim set of a patent. The claim set forms the legal heart of the patent and is well-structured, making statistical analysis feasible. The patent attorneys showed to be open for statistical patent valuation and saw several fields of applications within Philips, as for instance the benchmarking of own patent portfolios with that of competitors.

The interviews primarily answered the following question: *"what claim characteristics could hint you that a patent might be of higher-than-average value?"* The interviews delivered a set of quantifiable claim characteristics that will now be quickly summarized. The number of claims was believed to have a positive relation with patent value. This is because additional claims indicate more inventive contributions, they make successful prosecution les likely and they signal higher willingness of the applicant to invest in the patent. The length of independent claims was expected be negatively related to the scope of a patent, thus longer independent claims would limit the patent value. Independent claims could be divided into product and method claims. Basically, product claims protect a new product or substance, while a method claim protects the method of using a certain invention. Product claims have a wider legal protection and could therefore be of more value. Further, the way claims are formulated could be of influence on their value. If claims are formulated more functional, they might be of higher value than the

narrow-scoped so-called 'portrait' claims which describe an invention more by its looks. If a claim includes many 'limiting words' (which are believed to narrow the patent's scope), patent value was also expected to be lower. Finally, the ratio between the lengths of the preamble, which describes the state of the art a patent relates to, and characterizing portion, which describes the actual patented invention, could hint towards a certain patent value.

In addition, the patent attorneys stressed that differences in claim sets between applications and grants could exist. During the examination procedure, claim sets might change: some claims may be dropped or reformulated. This troubles the idea of early value prediction, which is based upon application data. Therefore, this thesis project specifically tested if application data could be used for value predictions. Including application data in the dataset also provided the opportunity to test if these differences could indicate patent value. One theory states that differences indicate a more narrow scope and thus lower patent value is expected. Another theory argues that differences are related to patents with higher expected value since they are normally filed with a much wider scope and thus more heavily debated during the examination procedure.

1.3 Empirical validation

Next, a dataset containing patents was constructed. The patents included in the dataset where earlier examined in a European research project that provided value estimations for them (Giuri et al, 2007). In total, this resulted in value estimations on about 2500 European (EP) patents in Philips-related technology areas. For these patents, the above mentioned claim characteristics were quantified for both grants and applications. In addition, a second set was built to test if time, technology field and country differences cause structural differences among claim sets. All three factors turned out to do so: American patents have on average more claims and shorter independent claims than their European counterparts; claim sets are influenced by the technology they describe; and throughout time, claim sets tend to become larger in volume.

The empirical validation started by analyzing granted claim sets. It was shown that claim sets generally consist of 1 to 20 claims. Most of them have one or two independent claims which have a length of about 150 words. Correlation was used to identify those claim characteristics that have a relation with patent value. Then, principal component analysis combined these characteristics into several components that were used in ordinal logit regressions. These regressions showed that a component related to the size of the claim set was the best significant predictor of patent value. When using basic claim characteristics, the number of (dependent) claims came closest to represent this component. If 'claim-set-size' characteristics were used together with three traditional patent indicators (forward citations, family size and opposition), they remained significant predictors. Therefore, the claim-set-size made a contribution to the existing indicators, although the explanatory power of the predicting models hardly increased by adding claim data. Overall, explanatory power was about 2.5% which is rather low.

The other, more sophisticated, claim characteristics did not show predictive power for patent value. The most surprising finding was that the length of independent claims could not predict patent value, although this was strongly expected by patent attorneys.

1.4 Early prediction and the differences between grants an applications

Next, application claim sets were used to see if claim data could indeed be used as an early predictor of patent value. Applications contain on average more claims and the length of their independent claims is generally shorter. 40% of all claim sets changed in the number of claims during the examination procedure while an even larger amount changed in their formulation (measured by changes in the number of words). The claim-set-size of applications had significant predicting power for patent value, although the explanatory power of the model

dropped to 1.6%. This confirmed the assumption that claim data can be used as an early value predictor. Further, the amount of differences between applications and grants, measured with the number of (dependent) claims, had a positive significant relation with patent value. A model based upon differences instead of the claim set size improved the explanatory power of the model by a 20%.

1.5 Conclusion and discussion

So, what is the contribution of this thesis project? First of all, it showed that the difference in patent claims between applications and grants is important although largely ignored by literature so far. Second, it answers with a firm 'yes' the question of whether full-text characteristics can relate to patent value. By looking at the size of the claim set, one can have, on average, an indication of patent value. However, this knowledge is not entirely new as earlier studies have showed similar results (e.g. Tong et al., 1994). Third, the report shows that the claim-set-size in applications is significantly related to patent value, suggesting that claim data could indeed be used for early value prediction of patents. These findings are not reported in earlier literature. Fourth, it is discovered that the amount of changes in claims during the examination procedure is positively related to patent value. This delivered a second value indicator based upon claim data which was not mentioned in literature before and provides a significant contribution to the existing indicators.

For many of the other claim characteristics mentioned in the interviews, a relation with patent value was not found. Still, these characteristics are believed to provide insight on the implicit rationales that patent attorneys use while analyzing or writing patent claims. To some extent, their identification and quantification can give researchers grip on the endogeneity (the interwoven relationship of claims with patent value) which Reitzig (2004a) mentions as the biggest disadvantage of full-text data.

The most important limitation of this thesis project with respect to early value recognition is that the current prediction model fails to take into account the fact that applications might change during examination or will not be granted at all. Backward-X-citations could perhaps be used for this, as they are positively correlated with the number of differences between applications and grants. However, this should be examined in future research as the currently used datasets were not capable of testing this. Future research could also focus upon a more accurate quantification of the identified claim characteristics so that they might relate to patent value in next studies. Finally, future research could look for other concepts that better fit the claim characteristics than patent value does, like perhaps the legal strength of the patent.

2 Preface

This report on my thesis project is not just a report on the relation between patent claims and patent value. It's an end product of many years of study. Education at the university is not just important to understand more about technology policy. For me, the primary task of a university is to change enthusiastic students into promising graduates. Promising because they not only managed to 'master' a certain topic, but also because they leave the university as analytically skilled individuals that formed an own opinion. For this, I can only be thankful to all who made it possible for me to experience these wonderful years at the Eindhoven University of Technology.

I was not able to write this thesis without the help of my supervisors. In all of them, I admire the enthusiasm and interest they always showed in the development of my project. I am most grateful for the huge amount of time they made for me, which gave me the opportunity to really dive into the topic. So, Piet, Gerard, Önder and Alessandro, thank you.

I was not able to finish my master degree without the help of my fellow students. Our group proved to be a group in which hard work was the norm. We pushed each other to our limits and therefore, I think we can all be proud of what we have achieved together. Needless to say that we had great times and many adventures together also! So, all my fellow students, thank you.

But most of all, I want to thank my mother and father. They have always shown interest in my education and stimulated me to do my best. I realize now that their sacrifices and support are not something 'normal': they can be proud of their selves for all they've done. Know that I am. Like in any good parents-child relationship, they not always favored my decisions, but nonetheless kept on supporting me. I am most grateful that they did. So, mom and dad, thank you for everything.

I hope the reader enjoys reading my thesis project. One time, I said that intellectual property is among the dullest subjects on Earth. Somehow I ended up in conducting my master thesis on them. Why? Well, once you've understood the basic principles….than you see what's really there: the latest technologies, a scientist's dream when it comes to available data, company strategies, legal fights, huge profits and so on. Intellectual property seems to be slowly reborn nowadays and I am happy in contributing a bit of knowledge in this exciting field.

Wouter

3 Contents

4 List of Tables

5 List of Figures

6 Introduction

6.1 The increasing importance of patents

Patents are legal documents that describe a certain invention. By obtaining one, its owner has the exclusive right to commercially exploit the invention the patent protects. This monopoly is granted to him by the national authority. As a return to society, the knowledge related to the creation of the invention is made public. Society gains since others can use this knowledge for further developments. Thus, the patent system aims to reward inventive activity while stimulating the accumulation of knowledge in society (Griliches, 1990, p. 1662-1663).

Patent law is a substantial part of Intellectual Property (IP), the generic term for copyright, patent law, trademarks and design law. In essence, IP always refers to a product of human thinking on which the ownership of that product belongs solely to the creator (Hoeth, 2007, p. 1-2). The role of IP became more important in recent decades. The number of patent applications rose drastically after the Second World War. Also in recent decades this growth continued steadily which is shown in the figure below for the number of European patents. Another indicator of the growing importance of patents worldwide is the establishment of the first international patent systems in the seventies. For instance, there was the establishment of the European Patent Convention in 1973, introducing a single filing route in Europe and laying the foundation of the European Patent Office (EPO) (Harhoff, 2004, p. 446).

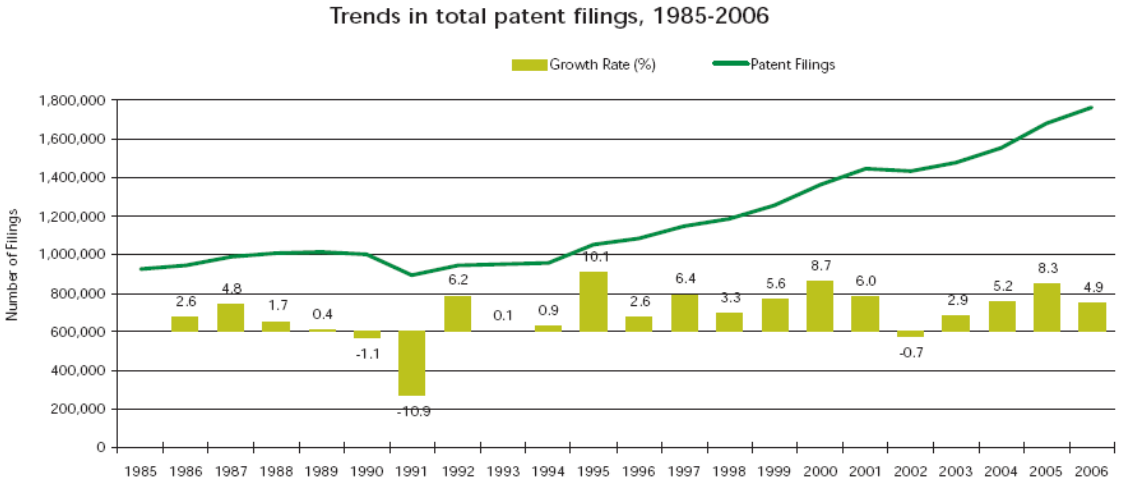

Figure 6-1 – Trends in total patent filings within the EPO from 1985 to 2006 - WIPO (2008), p. 14

Businesses adapted to the changing patent landscape. In the past, patents were used primarily as a defense mechanism to protect their own products. Nowadays, companies renewed their business models by actively exploiting their IP. In the scientific world, patents serve another purpose. Researchers embraced patents as a unique and valuable source of data available across a long time line. Especially in the field of economics, patents are highly regarded as being one of the scarce tools to measure technological development or innovativeness (Griliches, 1990).

6.2 Statistical Patent Valuation

Since patents protect a certain invention and provide knowledge to society, they have a certain value, like for instance a value related to their commercial potential. A key project of both researchers and businesses is to build

models that can predict this value. Currently, accurate value estimations require qualitative analyses by patent attorneys: engineers specialized in intellectual property law, having expert knowledge when it comes to patents. However, the qualitative analyses performed by them are time-consuming and expensive. Therefore, quantitative tools were developed in order to perform statistical patent valuation. Since patents are archived well-ordered in electronic databases, statistical analysis on their metadata is feasible. This metadata consists of information about the patent, as for instance the number of countries the patent is filed in or the number of years the patent is kept alive. So-called patent indicators are developed out of this metadata. They are tested on their ability to indicate patent value. Some of them are quite established as value predictors. The relation of other indicators with value is still examined and meanwhile even some not-yet analyzed metadata has been suggested as a possible value indicator (Zeebroeck, 2008).

Patent indicators are primarily useful when value assessments need to be made on large portfolios of patents, as the indicators lack the ability to give a reliable prediction for individual cases. A problem for most of the currently available indicators is that they suffer from timeliness: they are only available halfway a patent's lifetime. For business purposes, this is far too late. Practicing statistics on patents is considered 'tricky' sometimes, as patents are unique entities whose value is determined by factors that cannot be captured in metadata: its technical content, the size of the market the invention targets or the question whether or not the patent belongs to a standard. Therefore, it is essential that statistical analysis is combined with expert knowledge of the patent system to obtain correct results. The development of new indicators that can compensate for the problems that current patent indicators have is important if statistical validation models are to be improved. Therefore, this thesis project investigates the possibilities of a yet poorly explored piece of patent metadata: full-text patent data.

6.3 Reading Guide

This section will shortly describe the set-up of this report. The next chapter will review the literature. It starts with sketching the 'content and looks' of a patent. Next, the possibilities and drawbacks of statistical patent valuation are discussed and an overview of the most important patent indicators is provided. Finally, the literature review identifies the need for the development of an early available, full-text patent indicator.

The construction and empirical validation of such a full-text indicator was the goal of this thesis project. Therefore, the following research question was put forward in chapter 8: "*how can full-text patent data be used to predict the value of a patent?*" Also, three sub-research questions were formulated. Next, the research methodology is described. The methodology used in this project consists of two parts. First, a qualitative exploration on characteristics of full-text patent data that could indicate patent value was conducted by a series of interviews with patent attorneys. The interview findings are described in chapter 11.

Second, quantitative analysis was conducted to empirically test the relation of these characteristics with patent value. The construction of the dataset that was needed for this quantitative analysis is described in chapter 12. The chapter thereafter gives an overview of the characteristics of the variables in the dataset. Continuing, the actual data analysis takes place to provide an empirical validation of the identified text characteristics. The findings of the empirical validation are presented in three chapters. Chapter 14 examines which full-text characteristics can significantly predict patent value. The next chapter checks if these findings also hold when applications are used instead of grants (the exact difference between the two is explained later, but for now it is sufficient to say that applications are patents that have not yet been approved and are thus not 'granted'). The last chapter on empirical validation tests if the differences between applications and grants can also predict patent value.

At the end of this report, chapter 17 provides the major conclusions by answering the research question and the three sub-research questions. Finally, a discussion is presented to reflect on the findings in chapter 18.

7 Literature Review

The literature review starts with explaining what actually a patent is. Questions that will be answer are: what defines a patent, what is written in a patent, how can a patent be obtained, etc. Some of the different patent systems, filing routes and databases are discussed. This will all sketch the necessary background information. Next, an elaborate summary on statistical patent valuation and patent indicators is provided, including a review on their (dis)advantages. The chapter will then identify the need for a full-text patent indicator which is a promising, but so far rather unexplored, indicator. Then, an overview of present research findings on full-text patent analysis is given. Finally, the main challenges for the development of a full-text patent indicator are summarized.

7.1 Everything about patents in a nutshell

7.1.1 What is a patent?

A patent is a public document describing a certain invention and providing the owner of that patent with the exclusive rights for commercial exploitation of that invention (Markman et al., 2004 p. 536). A patent is granted by a governmental agency. Therefore, the patent system can be regarded as a trade-off between short-term social losses due to the granted monopoly in order to promote long-term technological development, which is eventually a gain for society (Griliches, 1990, p. 1662). A patent has a limited lifetime with a maximum of twenty years (in most countries). Patents may be used by their owners to protect products or production processes, obtain license fees from third parties, bargain cross-license agreements, block competitors, etc.

The requirements to obtain a patent on an invention are threefold. First, there is the 'novelty' criterion. It states that the invention may not be publicly available in other patents or any other kind of publication, both written and verbal, until the day of filing. This makes sure the invention is not part of 'the standard of the technology to that day' (Hoeth, 2007, p. 25-26). Second, the criterion of 'inventive step' says that only inventions described as 'non-obvious' by one skilled in that technological field should be allowed patent protection. This prevents the patenting of 'quasi-inventions'. Finally, inventions should be industrially applicable, which means that every invention should have a practical application (Festen-Hoff and Rijlaardsdam, 2004, 269-272). Despite examination on all three criteria, even a granted patent can still be annulled if, in a later moment, it fails to meet one of these requirements.

7.1.2 What does a patent consists of?

The basic ingredients of each patent are its *First Page,* the *Description* section and the *Claim* set. The *First Page* consists mostly of metadata: data about the patent, as for instance its filing date or the name of the applicant. This data is numbered according to the 'Internationally agreed Numbers for the Identification of Data' system (INID). INID codes enable metadata to be easily traceable in different patent systems or languages.

The first page shows among many other metadata a unique patent number and name, a filing and priority date of the patent, an overview of the IPC-classes to which the patent belongs and an overview of its references or citations. Other data included are an abstract of the content, the name of the inventor and name of the applicant of the patent. If a patent is filed through the EPO, it also shows its designated states (Golzio, 2006, p.4-5).

The second part of the patent is the *Description* section. It starts by sketching the related technology fields of the invention. After this, a summary of relevant prior art is given, presenting the 'current state of the art'. Typically, these parts introduce a certain problem: as a simplified example, this could be for instance the need to quickly

transport people over thousands of miles. The next step involves the introduction of the invention as the solution to these problems: in this example, the solution could be for instance a jet plane (provided it hadn't been invented yet) (Golzio, 2006, p. 5).

Third, there is the *Claim* set, which is the legal heart of the patent. It consists of a set of claims: single juridical-written sentences describing the invention that is claimed. The claims define the scope and boundaries of the patent protection. Using technical generalizations to describe the invention, applicants formulate claims as 'general' as possible, avoiding specifications where possible (Golzio, 2006, p. 5-6). This makes sure they claim as wide as possible, so that competitors will have more difficulties to work around their patent. The description section has a supporting function for the claim set: claims should be written in such a way that in case of possible misunderstanding of what the claims actually mean, the description section should clarify this. The formulation in claims differs greatly from normal language, as it is bound to certain rules and codifies technology using hard-to-understand jargon. This makes patent claims difficult to understand for those not skilled in the art.

7.1.3 Patent or patent family?

Patents have a national scope: they only protect an invention within country borders. However, together with technically equivalent counterparts abroad, they form so-called 'patent families' (Michels, 2001, p. 189-190). Several definitions of what constitutes a family are used, which could result in different family sizes. Mostly, this happens if different family members claim multiple priority dates. The most commonly used family definitions are the Derwent and the Inpadoc definition (it is beyond the scope of this paper to explain their exact definition). Family members can sometimes differ in their content since the outcome of prosecution is nationally bound. It may change the content of a patent in a particular country but not in other countries (Adams, 2006, p. 15-18).

In many articles on the development of patent indicators, only patents from a single country (mostly the USA) were used as unit of analysis. Also most patent indicators operate at the individual patent level. However, some indicators (like family size) can only be determined at the family level. Although the choice of analysis at the individual level is often one made for practical reasons, we must not forget the importance of the patent family. Scientifically, it is shown that the family level provides some advantages: there is a reduction of the 'home-advantage effect' (applicants tend to file in their native countries) and country-specific conditions are less influential (Criscuolo, 2006, p. 25). Finally, for business applications, family analysis is important as multi-nationals focus on a worldwide patent portfolio.

7.1.4 Patent systems

Patent filing is a national process which is done at the national patent offices. Well-known are the USPTO (United States Patent and Trademark Office), the JPO (Japan Patent Office) and the DPMA (Deutsches Patent- und MarkenAmt – German Patent and Trademark Office). The quality and thresholds of the examination procedures differ among countries. Partly, this is caused by differences in national patent law. For instance, the USPTO allows much more references as prior art since applicants are obliged to file all relevant prior art.

Because of the bureaucratic burden and high costs of national filing routes, two international patent systems have been established. One of them is the Patent Cooperation Treaty, in short PCT. Signed in 1970, this administrative system provides a PCT 'pamphlet' forming the patent application together with an international search report (Adams, 2006, p. 96-99). At that moment, the applicant has the choice to follow a Chapter 1 or a Chapter 2 PCT-procedure. The former means that the PCT pamphlet is sent to the national patent offices of the designated states, while the latter keeps the procedure at the international level by requesting a 'non-binding opinion' on patentability (Adams, 2006, p. 100-101). Once the application is sent to the national office(s), it is treated as it was

nationally filed. To prevent any confusion: the PCT procedure never examines nor grants, it only delivers a search report and, if requested, an opinion on patentability. Advantages of the PCT procedure are the extended time before one has to decide if one indeed wants to file a patent (18 months instead of 12 months), less translation costs and a search report which is accepted at the national offices (Adams, 2006, p. 96-99).

A second international system is in fact a European system: the European Patent Office, in short EPO. Founded by the European Patent Convention of 1973, the EPO examines applications on patentability and decides if they can be granted. However, EP grants are merely 'patents in no men's land'. They lack a territory where they can exercise the rights granted to them. In order to obtain these rights, an EP grant has to be sent to its designated states, where national offices then provide a national grant (Adams, 2006, p. 30-33). If an applicant wishes to file in four or more states, the EPO system becomes financially attractive (Grupp et al., 1999, p. 385).

7.1.5 Filing a patent

In the previous paragraph, three patent systems were presented. Each have an own filing route. The filing route is considered to be the procedure starting from filing an application and ending with the decision if a grant is awarded. There are many differences between the three routes and each route has many exceptions in itself. Nevertheless, some general lines can be identified. Obtaining a patent starts by filing a patent application to a national office, a PCT receiving office or the EPO. The date of filing is used as the priority date. After filing, applicants have one year counting from the priority date to extent their application to other countries or procedures. However, for the PCT, this period is 18 months (Zeebroeck et al., 2007, p.4).

18 months after filing, the application is published. This publication presents the content of the application which was hidden until then. In the ideal case, a search report joins the publication, judging if the patent meets the novelty criterion. However, due to capacity constraints, the search report tends to be published later in time. The search report gives an indication if the application will be granted without amendments: in case prior art was found, the application should be revised or can be even dropped.

After publication, the application is examined on its patentability requirements. In this phase, the claim set of the patent may still be altered by either the examiner or the applicant, as long as its scope is not widened anymore. The degree in which thorough examination is done differs among the patent offices. The USPTO has capacity problems causing their examination to be less strict than for instance in the EPO, leaving the actual decision on patentability ultimately to a judge in case the eventual grant is contested. Usually some years pass before an application is granted: for the EPO, the average time gap between the filing date and the date of granting is 4.2 years (Harhoff et al., 2004, p. 449).

7.1.6 Patent databases

All information on both applications and grants is stored in patent databases. Nowadays, patent offices offer this information freely in accessible online databases. Many large national offices like the USPTO provide an online search engine. EPO launched its Esp@cenet in 1999 which has become popular since (Adams, 2006, p. 153). However, these databases contain many errors and are not meant for extracting patent information on a large scale. Therefore high quality databases are established for professional users. Well-known is the Derwent World Patent Index of Thomson. This private database contains patent documentation from over 40 publishing institutions. Derwent uses a patent family classification system, rewrites titles and abstracts into an informative format and translates them to English if needed (Adams, 2006, p. 127-131). MicroPatent's database offers a search engine for patents, their family and their filing history, both on the level of bibliographic information and the full text level (Adams, 2006, p. 160).

7.2 Statistical Patent Valuation

The term 'statistical patent valuation' refers to statistical analysis using patents or patent families as its unit of analysis in order to estimate the 'value' of patents or patent families. Thus, patents are not qualitatively assessed in this process. Instead, they are regarded as equal entities differing in many metadata-characteristics: their number of claims, their priority year, their number of citations, the technology field they relate to, and so on. Statistical patent valuation will be further explained throughout this chapter, but one remark is important to make beforehand. That remark relates to the use of statistical patent valuation. The current features of statistical patent valuation are well summarized by Zeebroeck (2008). He states that statistical patent valuation could be used from the viewpoint of *"highlighting potentially more valuable patents from a huge mass"* (Zeebroeck, 2008., p. 20). However, as this field is still fully in development, statistical patent valuation can be expected to obtain more 'pinpointed' functionalities in the future. We should remind this when balancing between the benefits and problems regarding statistical patent valuation.

In the next paragraph, the fields of application of statistical patent valuation will be sketched to improve our understanding of statistical patent valuation. Thereafter, its advantages and disadvantages are discussed. Next, an overview is presented of the key patent indicators used in statistical patent valuation nowadays. Three remarks are presented regarding these indicators. These remarks introduce the need for a new sort of patent indicator: one that is based upon full-text analysis. But first, the rationales behind statistical patent valuation will be discussed.

7.2.1 The rationale behind statistical patent valuation

Before we continue, let us ask ourselves what is exactly the rationale behind using statistical patent valuation? There are two leading arguments which together form a solid base as to why statistical patent valuation is used. First, there is a steep increase in both the amount of patents applications and the size of patents. The growing IP importance is caused among many other factors by globalization, strategic patenting, emergence of new economies as South Korea and China and the growing role of new applicants as universities and small firms (Zeebroeck et al., 2007, p. 2). More patents bring more work for those entrusted with value estimations. Since qualitative analysis is very costly in both time and money, this forces businesses to look for new, more efficient ways of analyzing the patent landscape surrounding them. A second argument builds on the fact that the availability and accessibility of patent data is strongly improved. Combined with the increasing processing capacity to analyze such data, more methods of statistical analysis are feasible nowadays (Golzio, 2006, p. 4). Especially economists have shown interest and contributed by developing new indicators based upon patent metadata (Reitzig et al., 2004b, p. 2). Together, the need for new tools to analyze the patent landscape and the increasing possibilities to do so, provide the answer as to why the use of statistical patent valuation makes sense.

7.2.2 Fields of application

Knowing now why statistical patent valuation is used, the question pops up in what fields of application statistical patent valuation can be used. Already decades ago, scientists recognized the value of patents as information source to measure technological development or change-rates, innovativeness of firms or countries, etc. Here, patents are primarily regarded as an output indicator of R&D activities (Griliches, 1990, p. 1661). Later on, economists started using patents statistically to study patterns of industrial organization and technological spillovers (Reitzig 2004b, p.2) or to map technology trajectories. Further, the first business applications came into being. Competitive assessments and searches for profitable markets were undertaken with the use of patents by Ernst (1998). Breitzman and Mogee (2002) gave an overview of the many ways in which statistical patent valuation can be used to support intellectual property management, R&D management, technology assessment, but also in less-obvious fields as human resource management, company valuation and M&A assistance (Breitzman & Mogee,

2002, p. 187-202). The reader may have noticed that not all fields focus upon finding valuable patents. However, this thesis project will focus specifically on this functionality of statistical patent valuation.

7.2.3 Advantages and disadvantages of statistical patent valuation

In 1990, Griliches wrote about patents that *"nothing else comes close in the quantity of available data, accessibility, and the potential industrial, organizational and technological detail"*. According to his milestone article, there are plenty of options to extract information out of patents (Griliches, 1990, p. 1702). Ever since, many studies used the advantages of patent metadata for statistical analyses. The reasons for this are plentiful. Patents provide a detailed source of quality information and cover a wide range of almost all technologies (software excluded). There is a huge quantity of patents and all data is provided on a voluntarily basis (Hall et al., 2005, p. 21). The coverage of patents over a long period in time is a unique advantage, which is enhanced by their classification structure (the IPC codes) (Kürtössy, 2004, p. 95-96). Further, patents are one of the only objective measures of R&D activities. Finally, patents are one of the only sources for firms to timely monitor what their competitors are up to (Ernst, 1998, p. 280).

In the literature, also objections and warnings are expressed when it comes to statistical patent valuation. The main counterargument is that statistical analyses wrongly treat patents as a homogeneous group. Since patents are unique entities by nature, they form a very heterogeneous group and thus conducting statistics can be questionable (Griliches, 1990, p. 1666-1669). For instance, patents differ in their technical content, their scope of protection (e.g. single or multiple product protection) and have a highly skewed value distribution (Kürtössy, 2004, p. 96; Simmons, 1991, p. 33-34, Giuri et al., 2008, p. 1121). Further, patents can structurally differ due to discrepancies in law across countries. Michel and Bettels (2001) show that the origin of the patent office examining a patent (EPO, US or JPO) influences the quality of the search report and the number of citations. Finally, differences between industrial sectors or firms can cause structural variations (Kürtössy, 2004, p. 96).

The main point that should be taken here, however, is not that using statistical patent valuation is wrong. Instead, it can be used, but one should do so with caution. One should be aware of the limits and biases of using patents as a data source, which requires knowledge of the patent system. Perhaps this is best formulated in the following statement made by the economist Jacob Schmookler: *"'We can choose whether we wish to use patent statistics with prudence and to learn what we can learn from them, or not to use them and to do without all the information that they alone can provide"* (Fabry et al., 2006, p. 216).

7.3 Patent Indicators

There are many metadata characteristics of patents that are analyzed in statistical patent valuation. Literature speaks of the construction of so-called 'patent indicators' which refers to the ability of certain characteristics in patents that can indicate certain 'phenomena'. In most cases, patent value is indicated, but the field of applications is much wider as was shown in section7.2.2. In this chapter, the most commonly used patent indicators will be discussed briefly[1]. What could be confusing is that some indicators were used as independent variables testing their relatedness with value, while other studies used the same indicators as a dependent variable representing patent value. This may sound strange, but is explainable by the fact that some indicators became established predictors of value over time and therefore were used as dependent variables in later studies.

[1] For an elaborate summary of each of these characteristics or indicators, one can consult the appendix. In addition, the reader is advised to look into the paper by Zeebroeck and Pottelsberghe de la Potterie of 2007 named 'Filing Strategies and Patent Value'.

7.3.1 An overview of the most important patent indicators

The most established value indicator is forward citations: the amount of references a patent receives over time. The more a patent is cited, the more central its role is within that technology field. Thus, a higher patent value is expected (Trajtenberg, 1990; Michel et al., 2001). Another well-established indicator is litigation and/or opposition. The mere fact that a patent is disputed in court or by the EPO-opposition procedure signals third party interest and the willingness to invest in the necessary legal steps. This is an indication of value (Lanjouw et al., 2001; Harhoff et al., 2004). The amount of times renewal rates have been paid is another established patent indicator. It tells us how long a patent is kept alive by its owner. Since the price of renewal rates increases over time, those patents that are kept alive longer are expected to be more valuable (Pakes et al., 1984; Zeebroeck et al., 2007, p. 6). A fourth indicator would be family size: the number of countries a patent family covers. Since each designated state brings additional costs for filing and renewal fees, larger families are expected to be more valuable (Putnam, 1996).

Backward citations have a more blurry relation with value than the indicators mentioned so far. Representing the amount of references a patent gives to other patents or non-patent literature (as for instance scientific articles), backward citations seem to be negatively related to patent scope and thus value. However, findings indicating the opposite are also available and therefore, backward citations are no established value indicator at the moment (Harhoff et al., 2003). A relatively new indicator is the filing strategy that was chosen by the applicant to obtain the patent. The EPO and PCT procedures offer several routes of filing that can lengthen or shorten the application process. The choice of a particular route can thus be considered a tactical choice of the applicant, which could be related to (expected) patent value (Reitzig, 2004a). Some researchers relate applicant and/or owner characteristics to patent value. For applicants, the size of the institution they work for and experience they have with the patent system can signal value, whereas for owner characteristics, company value or Tobin's q can be related to patent value (Giuri et al., 2007). Finally, in some cases several of the indicators above were combined to create a composite index. Since indicators differ in the aspects of value they predict, composite indices have an advantage here, although they are difficult to construct (Zeebroeck, 2008).

7.3.2 Three remarks on the traditional patent indicators

This section provides the reader with three remarks concerning the 'traditional' indicators described above. The remarks will focus upon the need for indicators with a rationale based upon social rather than monetary value, the fact that traditional indicators fail to take into account full-text patent data and the fact that most of the traditional indicators become available late in patent's lifetime.

So, firstly, what dimension of value do traditional patent indicators actually indicate? Briefly, two value concepts can be identified: the monetary value of a patent and its social value. The monetary value or asset value relates to the profits that were gained by the patent owner (Heiden, 2001, p. 7). The social value relates to the contribution of the patent to society. It is assessed by examining the impact of the invention protected by the patent and/or the knowledge revealed by the patent[2] (Heiden, 2001, p. 8). In a slightly different fashion, Reitzig (2004a) also describes these two value concepts. He describes 'observable' effects of a patent (the influence on prices, costs and sold quantities of products by patent protection) and 'unobservable' effects. These 'unobservable' effects relate to latent value determinants: the scope or breadth of a patent, the 'size' of its inventive step, the degree of novelty and the state of the art a patent relates to are some of these determinants (Reitzig, 2004a, p. 940).

[2] For a more elaborate discussion on the topic, the reader is referred to appendix II

To my opinion, some of these unobservable effects (like the degree of novelty, its scope or the 'size' of its inventive step) relate more to the social value a patent has than to the monetary value. I believe this, because these latent determinants relate to the knowledge included in that patent. If a patent scores high on these determinants, by being e.g. highly innovative and having a broad scope (which can signal a large amount of practical applications), high social value can be expected in the long run. Continuing this line of reasoning, let us assume the ideal situation in which the patent owner can fully benefit of all value given to society by the patent. In other words, the patent owner maximizes its monopolistic advantages. Here, we are talking about the monetary value of the patent. Now, the observable effects become a measure for the degree to which the patent owner succeeded in exploiting the potential of the unobservable effects. In this sense, I believe the observable effects can be related to monetary value of the patent.

Looking again at the traditional patent indicators, one should conclude that most current indicators are based on a rationale that makes use of the observable effects. The rational of using family size, renewal rates and litigation and/or opposition is rooted in monetary considerations, as it is also the case for most of the less established value indicators. This does not mean that these indicators have no relation with the latent determinants: a highly opposed patent is very likely to score high on the determinant "difficult to invent around". The point made here is merely that it would be interesting to develop indicators that were based upon a rationale related to the social value of a patent. To some extent, this is the case for forward citations, as some scholars claim that forward citations are more likely to indicate social value than monetary value (for instance, Trajtenberg, 1990, p. 178). For now, the question of which concept of value is measured will be abandoned. The question is worth a study in itself[3] and will not be answered here. The point that should be taken here is that there is a need for an indicator measuring the intrinsic quality of the content of a patent and is thus based upon the social value of a patent. However, note that this statement is personal and not backed up explicitly by literature.

A second remark about the traditional patent indicators is that all of the indicators use so-called 'front-page' patent information. They fail to make use of information contained in the patent itself. Experts as patent attorneys are therefore suspicious about such front-page indicators, since "*most of the information on a protected technology and its anticipated economic value is conveyed in the patent draft itself*" (Reitzig, 2004a, quoted from p. 942). According to Reitzig, 'first-page' indicators belong to the category of 'first' and 'second' generation indicators, although in my view, the difference between them is kept rather vague. However, both generations are using 'front-page' patent metadata and differ in this from the 'third' generation indicators that characterize themselves by using the full-text of a patent. Reitzig states these third generation indicators have a greater potential than the first and second generation and therefore should be further developed (Reitzig, 2004a, p. 942-943).

A last remark is made specifically about the problem of timeliness that troubles most of the established value indicators (forward citations, litigation and / or opposition, renewal rates and family size). Most of these indicators become available at a late stage in the patent's life. Forward citations are believed to become reliable somewhere between 4 and 5 years after filing. Renewal rates are useful somewhere between 4 to 10 years after filing. Finally, litigation or opposition outcomes are expected at earliest at 4 years after filing (Reitzig, 2004a, p. 940; Van Zeebroeck et al., 2008, p. 4-5 & 9). Only family size has some ability to early indicate patent value. The moment the applicant has to decide which state it designates is the first moment a (preliminary) family size can be determined. For national or EPO filings, this moment is one year after the filing date, for PCT filing it is at 18 months after filing.

[3] See the article of Zeebroeck of 2008 named 'The Puzzle of Patent Indicators'

To conclude, since three out of four of the established indicators are available rather late in a patent's lifetime, there is clearly a need for value indicators that are available more early.

7.4 A full-text indicator

The format of a patent is standardized to a large extent and allows an easy storage and recovery of both the patent itself and all sorts of metadata. For this and several other reasons, patents provide an excellent source of data. However, as was described above, most traditional indicators based upon this data only make use of the 'front page' information of the patent. They fail to relate to the invention itself or to relate to the intrinsic quality of patents, described by the latent determinants mentioned by Reitzig (2004a). In addition, most established indicators suffer from timeliness: they become available late during a patent's life. Therefore, it would be interesting to explore new metadata of patents that are content-related and are available early in a patents lifetime. Having said this, it may come as no surprise to the reader that this paper aims to use full-text patent data as a value indicator. In the next paragraph, an overview is given of the findings of earlier studies on full-text patent analysis. Further, it is explained why such studies require a specific research environment with experts that have an excellent understanding of how to read and write a patent.

7.4.1 Previous studies on full-text patent analysis

In most studies on full-text patent analysis, the text in patent claims was used. According to literature, the claims section describes the actual scope of protection, or, as Markman et al. (2004) puts it, "*analogous to the metes and bounds of a deeded property, claims define the scope of an invention and distinguish its property from the surrounding technological territory"* (Markman et al., 2004, quoted from p. 536). EPO Art. 84 states that "*a fair statement of claim is one which is not so broad that it goes beyond the invention nor yet so narrow as to deprive the applicant of a just reward for the disclosure of the invention"* (EPO, 2007b). Thus, it can be expected that applicants always aim to write a scope balancing precisely on the edge of what would be accepted by an examiner. Further, applicants wish to protect those inventions with high expected value as much as possible. One way to do so is by including additional claims to create more fall-back positions so that successful prosecution of the patent becomes more difficult (Gambardella et al., 2006, p. 10).

A first study in the field of claim analysis was done by Tong and Frame in 1994, who showed a positive correlation between the number of patent claims and technological performance on an international level. The assumption was that each additional claim represents a *'distinct inventive contribution'* (Tong et al., 1992, p. 134). Hence, more claims would signal an invention with a larger field of applications and thus it could indicate higher patent value. Second, they proved that analyzing claims provided a better indication of inventiveness than merely counting patents. In later studies, claims were related with forward citations, backward citations and litigation (Lanjouw et al., 2001, p. 140), family size and renewal date (Zeebroeck et al., 2007, p. 22) and opposition. However, the relation with opposition was probably caused by the fact that high numbers of claims create varying expectations about the actual scope of the patent. More claims make it more difficult to determine the exact scope of the claim set, in which case opposition is more likely (Harhoff et al., 2004, p. 461). Reitzig (2004a) showed a significant correlation between the number of independent and dependent claims and the likelihood of opposition. In addition, Reitzig tried to relate the number of words describing the state of the art and the technical problem, which stem from the description section of a patent, with the likelihood of opposition. However, he found only weak evidence for this relation (Reitzig, 2004a, p. 953).

Several problems using patent claims are identified by literature. The number of claims differs per country. Further, there is an overall increase in the amount of claims per patent over time (Tong et al., 1994, p. 133-138). Zeebroeck et al. add the problem of differences in the number of claims among technology fields (Zeebroeck et al.,

2007, p. 12). Also, the length of claims is technology field dependent. Since claims also define the state of the art, more mature technologies would have longer claims by nature, making cross-technology-comparisons difficult (Gambardella et al., 2006, p. 10).

7.4.2 Finding the right research environment

Examining full-text patent data is difficult because of two reasons. First, patents are written in a 'legal language'. The patent world has a unique jargon and ways of describing inventions, which is related to the fact that patents should serve multiple purposes: (1) they should describe an invention into detail and (2) they should have a scope as broad as possible and (3) they should uphold themselves in case of litigation. Knowing exactly how text in a patent can be used as a value indicator requires an understanding of what features, characteristics and/or tools patent attorneys use in their strive to formulate patents as profit-making as possible.

A second problem related to full-text analysis according to Reitzig is the endogeneity of patent claims in their relation to patent value (Reitzig, 2004a, p. 943). This endogeneity should be seen in the following way. The way a claim is drafted determines to some extent the value of the patent. But at the same time, claim sets have been drafted in a certain way since a certain patent value was expected. Thus, those patents promising to become important might be formulated with much more attention and thus higher quality, and patents of higher quality are eventually better equipped to become of higher value. In order to entangle this relation, a strategy would be to get grip on the characteristics of claims that indicate patent value.

For these reasons, a study on a full-text patent indicator has to be conducted in a research environment where it is possible to obtain expert knowledge on patents. In-depth interviews with patent attorneys is the most likely way of extracting this kind of information. Their (tacit) knowledge is crucial when it comes to understanding what characteristics in the text of a patent can hint towards expected patent value. Therefore, an environment is required that employs at least several of them. In addition, this environment should provide access to quality patent databases. This is needed to extract text data in an orderly format so that full-text indicators can be constructed in order to empirically validate this new indicator in a later stage.

8 Research Question

In the literature review, the need for a patent indicator making use of full-text patent data was identified. Therefore, the general research question will be:

How can full-text patent data be used to predict the value of a patent?

Answering this question will be done in two steps. First, characteristics of the text in a patent that has a possible relation with patent value have to be identified. Second, by statistically testing the relation of these characteristics with patent value, a prediction model can be constructed. The functioning of this prediction model can then be compared with the traditional indicators. In particular, the statement that full-text data can perform as an early indicator will be tested.

The following sub-research questions are defined:

1) **Which characteristics of the full-text of a patent could be related with value? And what external factors may disturb these relations?**

Answers on these questions require interviews with experts skilled in writing, reading and understanding the text of patents: patent attorneys. By interviewing them, several characteristics could be identified that could be related to patent value. The second part of this sub-research question makes sure that external factors that could influence this relation will be identified.

2) **Which of the identified characteristics are significantly related to patent value?**

The characteristics identified by the patent attorneys will be statistically tested on their ability to indicate patent value. In order to do this, all characteristics will be related with a value indicator while controlling for possible external factors.

3) **To what extent does a full-text indicator improve the statistical methods to analyze patent value?**

A full-text indicator will be constructed out of those characteristics that best predict patent value. This indicator is tested against several of the traditional indicators in order to see what it contributes. In addition, the argument that full-text analysis provides a value indication early in a patents lifetime will be tested.

9 Methodology

This thesis project aims to combine qualitative and quantitative research. It started with a broad literature review, giving a general introduction into patents, after which the role of patent statistics and the use of indicators has been discussed thoroughly. It ended with the identification of the need for a full-text patent indicator. The main research question focuses on first identifying characteristics in the text of a patent that can indicate or even predict patent value and then statistically test them.

In order to answer this question, a research environment is needed in which it is possible to conduct in-depth interviews with patent attorneys. In addition, this environment should provide access to quality patent databases so that a dataset can be constructed for the statistical analysis.

Fortunately, the department of Intellectual Property and Standards (IP&S) of Philips was willing to participate. Being the company that employs the highest number of patent attorneys in The Netherlands, IP&S is the ideal environment to conduct interviews in order to explore the relation between the text of a patent and patent value. In addition, via the sub-department of IP&S called Business Intelligence, access to quality patent databases was obtained. Moreover, this sub-department uses statistical patent valuation itself. For this, it developed a tool called PLATO which uses several established patent indicators. Both the access to the databases and the ability to make use of the PLATO system makes statistical analysis possible.

The following steps in research are defined. First, several interviews will be conducted with patent attorneys to identify a spectrum of text characteristics and their possible relation with patent value. More details on the way the interview sessions are set up can be found in chapter 11. Second, a dataset of patents will be constructed consisting of A) quantified text characteristics, B) value indicators/estimations and C) influencing external factors. Quantifying text characteristics will be done by transforming characteristics of the patent text into numerical values. This patent text will be extracted from quality databases for a set of patents of around 2700 patents. This is a subset of a set of around 10.000 EP patents that received a value estimation in a broad European study on patent value[4]. The choice was based upon relatedness to the technological fields in which Philips is active. Besides these value estimations, several other value indicators will be constructed or extracted from the database or the PLATO tool. Finally, data on external factors will come from the databases. More information on the exact construction of the dataset can be found in chapter 12.

The third step in this research project will be a statistical analysis for empirical validation of the relation between the identified characteristics and patent value. As a first step, the data is explored and possible external factors that could influence the findings are tested. Those factors that turn out to be relevant will be included in later statistical models in order to control for them. Next, all characteristics will be examined on their relation to value, after which a selection of 'best predictors' will be used to construct a full-text indicator. This new indicator will be tested against several indicators already used to measure its contribution. Then, an assessment will be made on the ability of a text indicator to predict patent value early in the patents lifetime. This is done by using application data. To finish off, the difference between applications and grants is tested on its relation with patent value.

The last two chapters of this paper consist of a conclusion and a discussion. In the conclusion, the (sub-)research question(s) will be answered. Finally, the discussion will reflect on these findings.

[4] See Giuri et al., 2008, p. 1107-1127

10 Patents within Philips

Before we continue with the section on the interviews with patent attorneys, first a brief introduction is given into IP&S, the sub-department Business Intelligence and the PLATO tool.

10.1 Philips IP&S

In a response to the growing importance of IP, most large companies nowadays have a specialized IP department. Where patents were used in the past mostly as a defensive mechanism, companies are now actively looking for licensees, set up standardization projects or trade in patents. Their challenge is to exploit their intellectual assets as much as they can. New business models are developed to cope with the increasing amount of patents.

One of the companies active in this field is Philips, the famous multinational we all know for developing and manufacturing technical products in the fields of healthcare, lifestyle and lighting. The company employs around 128.000 people in over 60 countries. It has an annual turnover of 27 billion euro's. Philips is active in the field of intellectual property rights: it holds around 80.000 patents, almost 30.000 trademarks and has almost 45.000 design registrations. The department of Intellectual Property and Standards (IP&S) is the central organ for all intellectual property related issues within Philips. In 26 offices spread all over the world, its 450 employees focus on making "*IP&S a leading industrial IP organization providing world class IP solutions to Philips' businesses to support their growth and competitiveness"* (Vision of Philips, IP&S). IP&S is headquartered at the High Tech Campus in Eindhoven, The Netherlands.

Patent attorneys form the largest group of employees within IP&S. They are educated in both juridical and technical matters. Their main task is to write patents, advice in licensing or standardization matters and keeping a close eye on the Philips patent portfolio. Patent attorneys are mostly looking at patents from an individual, detailed and qualitative level. Generally they can be seen as the patent experts.

10.2 Business Intelligence

One sub-department within IP&S is dedicated to information delivery on patents: Business Intelligence (BI). Their job is to provide patent attorneys and IP&S management with the information they need, which is summarized in their mission statement: "*to contribute to the IP&S processes with reliable high quality market and IP information and analysis"*. One of their primary tasks is to create 'patent landscapes': overviews of relevant sets of patents in a certain technology field or of a particular company.

Starting point in the creation of a landscape is a request from business, for example for an overview of all patents of company X in the field of medical imaging. The relevant patents are gathered, which requires expert knowledge of and experience within the field and results. Second, an overview of these patents with some (graphical) analyses is constructed. Landscapes can be technology orientated, presenting an overview of the patents in a certain technology field, or company orientated, giving insights in the portfolio of a particular company. Either way, besides giving an overview of which patents are relevant, patent landscapes aim to sketch the relations, trends and patterns in the landscape.

10.3 PLATO

In recent years, BI developed a system to enhance their landscapes: PLATO, which stands for the Patent Landscape Analysis Tool. It uses patent metadata to provide additional information and insights. PLATO uses a ranking mechanism that operates by giving each patent family in a landscape a certain value based upon metadata indicators, as for instance family size or opposition.

After the relevant patents for a landscape have been selected, they are uploaded in PLATO together with numerous metadata[5] of the patents. Now, PLATO is capable of exporting all patents and their metadata in family-orientated Excel-format. The added value of this format is that it makes it easier to assess a patent portfolio. In addition, PLATO gives aggregated scores on the family level for several patent indicators (family size and two related indicators at regional levels, forward and backward citations and opposition). Ultimately, it even gives an 'Overall Score' combining all indicators into a weighted average representing the 'quality' of the patent family. This composite index is used within IP&S as a guide to determine which patents are more or less interesting in a landscape.

PLATO is developed grass-roots by combining scientific literature with practical knowledge based upon experience. This also holds for the ranking mechanism, which has been only marginally validated. Thus, it is not clear what 'quality' the Overall Score actually indicates.

Now that IP&S is introduced in this chapter, the next chapter will continue with the findings of a series of interviews. These are held with patent attorneys working at IP&S in order to explore the possibilities of developing a full-text patent indicator. Interesting is that in the early development days of PLATO, an indicator using patent claims was proposed for development and implementation in the system. Unfortunately, other indicators proved more promising at that moment and this plan was rejected. However, BI has shown interest to develop and implement a full-text indicator based upon the findings of this project, if its contribution can be shown.

[5] Metadata as e.g. priority year, citations, IPC classification, applicant, inventor, legal status etc.

11 Expert Interviews

This chapter will present a summary of expert interviews held within IP&S. First, the interviewees and the general set-up of the interview sessions are presented. Second, the identification of several text characteristics is summarized. Third, the opinion of the experts on the use of statistical patent valuation is discussed. Finally, some reactions from patent attorneys on the 2004-article by Reitzig are presented, since the article has a close relation to this thesis' topic.

11.1 Set-up of the interview sessions

There were in fact two interviews rounds. At first, an exploratory round of four interview sessions was held with four patent attorneys with the function of IP Counsel. These interviews were conducted in order to determine how full-text analysis can be used to determine patent value. The four interviewees agreed on the fact that out of the entire patent, its claim set is the most important textual part to analyze. The claim set forms the heart of the patent and describes what is actually claimed, while its description part has merely an explaining and clarifying role. Therefore, it has been decided that the second round of interviews would be used to explore the characteristics of a claim set has. In all interviews, the guiding question was: *"what claim characteristics could hint you that a patent might be of higher-than-average value?"* In addition, the patent attorneys were asked for their opinions on statistical patent valuation and to comment on some statements of the 2004-article by Reitzig.

In the second interview round, a total of 11 patent attorneys were interviewed. Eight of them also have the function of IP Counsel within IP&S. Two of them were selected because they have an American background and had experience in dealing with US patents and the USPTO. This gave access to knowledge about the US system which was not obtainable from the other six Dutch IP Counsels. The other three interviewees work at the management level of IP&S. Two of them have the function of IP Portfolio Manager and the other is an IP Counseling Manager.

The interviews can be considered as semi-structured. Several questions were prepared beforehand but a strict question-answer structure was not used. Input from earlier interviews was used in later interviews, so that several opinions on claim characteristics were gathered. All interviews were conducted by myself.

11.2 Claim Characteristics

This paragraph will summarize the claim characteristics that were identified in the interviews and their relationship with value. Background information on these characteristics was also taken from the EPO Guidelines of December 2007 (EPO, 2007b). At the end of each description of a characteristic, a hypothesis is presented giving the expected relation to patent value.

11.2.1 Number of Claims

Scientific studies already showed the relation between patent value and the number of claims, based upon the idea that each additional claim represents a *'distinct inventive contribution'* (Tong et al., 1992, p. 134). The patent attorneys confirmed these findings qualitatively, with the assumption that the examiner did a good job in all cases. The interviews also provided an additional reason why claims relate to value. It turned out that applicants pay extra for larger claim sets. In the EPO, until 2008 an additional fee of $\pm$ 50 euro was required for the eleventh and each subsequent claim. In 2008, policy changed with even higher fees: from the sixteenth onwards each additional claim costs the applicant 200 euro (EPO, 2007a; EPO, 2008a). The USPTO charges applicants for each fourth and subsequent independent claim little over 200 US dollars while each dependent claim in excess of 20 will cost 50 US

dollars (USPTO, 2009). These costs might seem low for multinationals, but since they tend to file many patents, they can still be considerable for them. Therefore, the willingness of applicants to pay the extra fees indicates a higher expected patent value.[6] Since the only drastic change in fee height policy was very recently (going from 50 to 200 euro's) it is expected that this policy is time-independent until 2008.

The attorneys mentioned that the amount of claims depends heavily on technology field. For instance, the amount increases quickly when looking into technologies with cooperating technological constructions. Also there is a country dependency. Unlike in the EPO, the USPTO occasionally allows patents with extreme amounts of claims (numbers over a hundred). According to one of the American attorneys, this is done to make prosecution against the patent more difficult. For these reasons, the number of claims should be analyzed in a group as homogeneous as possible.

Based upon these findings, the following hypothesis was put forward.

- **There is a positive relation between the amount of claims in a patent and the value of that patent**

11.2.2 The length of claims

As a joke, attorneys sometimes use the rule of thumb that every claim exceeding ten centimeters in length is a poor claim. What is meant is that ceteris paribus, longer claims (thus claims containing more words) limit the scope of the patent. The more elements in a claim, the more limited its scope is. Since additional elements require additional words, fewer words could indicate a wider scope. Further, longer claims make it more difficult to force patent rights, partly because the content of the claim becomes less clear.

Claim length is believed to be dependent of several external factors. The maturity of the technical field plays a role, since a more mature field will require longer claims because of extra prior art. Counterargument here is that mature fields have established jargon and definitions can shorten claims. For instance, what is now called a compact disc was called a 'recording disc for audio and video' or a 'reflective optical record carrier'. Second, length of claims differs among technology fields. Some technologies simply require longer claims by nature. Third, length of claims is dependent upon the writing style and skill of the applicant. Fourth, different languages are intrinsically shorter or longer. This means that absolute comparisons can be only made within the same language.

When looking only at applications, short claims could also indicate bluff of the applicant, or a high level of understanding of what the applicant wants to protect. Only when the grant is provided it becomes visible what exactly was the case.

For all these reasons, the following two hypotheses are put forward.

- **There is a negative relation between the length of a claim in a patent and the value of that patent**

- **There is a negative relation between the length of the first claim in a patent and the value of that patent.**

(The first claim of a patent is in most cases thé most important claim. The hypothesis is added to see if looking only at the first claim can provide significant results, which would drastically decrease the amount of data that has to be analyzed)

[6] Additional study pointed out that Zeebroeck et al. (2007) had also discovered this. (Zeebroeck et al., 2007, p. 13)

11.2.3 Independent and Dependent Claims

All claims are either independent or dependent. Independent claims are the core of the claim set. Each independent claim in a set literally 'claims' the invention the patent is trying to protect. Of course, these claims are defined as broad as possible, to widen the scope of the patent. Multiple independent claims can be included in one patent for two reasons. First, it is allowed to have an independent claim in different 'categories' (these categories will be in section 11.2.4). In the US, it is even allowed to have several independent claims per category. Second, certain technologies as for instance transmitter-receiver systems or plug-and-socket applications are allowed to have independent claims for both parts of the system (EPO, 2007b, paragraph 3.2). Further, differences in law exist between the strictness of allowing several independent claims. While the EPO is very strict with pushing the applicants towards filing only a few all-covering independent claims, the USPTO is more loosely in this.

Most of the time, an independent claim is followed by several dependent claims. A dependent claim includes all the features mentioned in the independent claim it refers to and contains 'particular embodiments' in addition. Therefore, dependent claims have the function of signaling (and claiming) specific uses of the invention as well as the function of fall-back positions. The reference of the dependent claim towards the independent claim should be included at the beginning of the dependent claim, if possible. In some cases, also independent claims can refer to other claims; these references are found mostly at the end of the claim (EPO, 2007b, paragraph 3.4).

The link with patent value was related mostly to independent claim. Two sorts of relations were suggested. First, a larger amount of independent claims could signal value. The rationale is that if the applicant pushed for more independent claims despite the troubles it gives, he must expected high value. Second, the length of these claims is supposed to be negatively related to patent value for the same reasons as were mentioned in the paragraph on claim length.

For independent and dependent claims no hypothesis is given. Instead, the point that should be taken from this paragraph is that since independent claims are the base of the claim set, most characteristics will be analyzed only for independent claims. Of course, where the difference makes sense (as for instance in counting the number of claims), both independent and dependent claims will be analyzed.

11.2.4 Product and Method Claims

In general most independent claims are either product claims or method claims. A product claim describes the features of a substance or physical entity, while a method claim describes the use of such a substance or physical entity. Product claims give a wider protection range than method claims and are thus expected to be of a higher value. Within product claims, one can subdivide apparatus claims, system claims and device claims, but basically every noun as the base of a claim could be a product claim; e.g. 'an airplane' or 'a fiber'. In the US, these claims are also called 'composition of matter' claims. Method claims can also be formulated as 'process' claims. Mostly, a method claim is filed together with a product claim. However, in some cases already patented or known inventions are used in such a novel fashion that this can be considered as an invention on itself. In such cases, product claims cannot be used and therefore also patents exist consisting only of method claims.

So, in the end we are interested in answering several questions. One of them is to examine if product and method claims can predict patent value. Another is if there is a difference in value between patents containing product claims and patents containing method claims.

Therefore, the following three hypotheses are put forward.

- **There is a positive relation between the amount of product and / or method claims in a patent and the value of that patent**
- **There is a negative relation between the length of product and / or method claims in a patent and the value of that patent**
- **Patents containing product claims have a higher value than patents containing method claims.**

(It is difficult to pinpoint what relation to value can be expected for method claims. Attorneys preferred product claims over method claims but in method claims are still good claims. Therefore, in the end one can expect a positive relation between the amount of method claims and value. For the length the same reasoning holds as for product claims)

11.2.5 Portrait versus Functional Claims

Although not an official term in patent law, attorneys mentioned that all claims are formulated either more 'functional' or more in a 'portrait' shaped way. When a claim is literally describing the construction of an invention, it is creating a sort of 'portrait' of that invention. This is called a portrait claim. The opposite of a portrait claim is a functional claim. This is a claim describing the functionality of an invention. For example, a door could either be described by 'a wooden rectangular object fitting in a doorpost' or by a 'means of separating two spaces': the former is a portrait claim, the latter a functional one. The reader should notice that the latter covers a much wider area, be it simply by the fact that the shape of the door does not to be rectangular or the door does not need to be made of wood.

Portrait claims have a narrow scope. Anyone willing to use the technology protected has to alter little in order to avoid falling under its protection. The use of portrait claims could indicate that the applicant was in a hurry or failed to think of alternative uses of the invention. However, sometimes prior art can leave the inventor with no other choice. The scope of functional claims is thus much wider which makes it much harder for competitors to work their way around a certain functionality. Therefore, functional claims are regarded as more valuable. Functionality can be identified to some extent by the phrase "for ….ing", like for instance in the following claim: *'A patient positioning system for positioning a patient undergoing radiation treatment….(etc)'*. (patent number EP1702650 filed by Hitachi, Ltd.) Portrait claims were not believed to be identifiable with a simple rule.

An example of the first independent claim of patent number EP1867584. Also the second claim, being a dependent claim, is given. The patent is claiming the invention of coffee pads.
….claimed is:
1. Coffee bag comprising
• a first and second sheet, which are joined while overlapping at the circumferential edges so as to form a closed envelope of an inner space situated between the first and second sheet;
• a coffee bed with ground coffee that is present in the inner space;
wherein the first and second sheet are waterpermeable but not permeable to the ground coffee, characterized in that said envelope is provided with a functional ingredient.
2. Coffee bag according to Claim 1, wherein the functional ingredient is provided on the outside of the envelope.

On the functionality of claims, the following hypotheses were put forward:

- **There is a positive relation between the amount of functional claims in a patent and the value of that patent**
- **There is a negative relation between the length of functional claims in a patent and the value of that patent**

Instead of identifying if a claim is functional, one can also count the number of functional phrases identified. It is assumed that the more functional phrases are used, ceteris paribus, the more functional a claim is. Therefore, it is expected that this will have a positive relationship with value. The hypothesis in this case is:

- **There is a positive relation between the amount of functional phrases in the claim set of a patent and the value of that patent**

11.2.6 The Preamble and the Characterizing Portion

The EPO uses a two-part form structure for independent claims. This means that claims should be formulated in two parts. The first part, the 'preamble', should describe all relevant prior art needed to define the invention. The second part, the 'characterizing portion', describes the 'new' features that the invention adds to the prior art (EPO, 2007b, paragraph 2.2). During examination, it can happen that features that are thought not to be novel by the examiner are transferred from the characterizing portion to the preamble. Use of the two-part form is not obliged and is technology field dependent: some technologies fit better into the model than others. The end of the preamble and thus starting point of the characterizing portion can be identified in most cases by the phrase "characterizing".

It has already been discussed that claim length can indicate patent value. Therefore, it was suggested to the patent attorneys that perhaps the length of the preamble and / or characterizing portion, or their ratio, could have a relation with patent value. The attorneys agreed that a short preamble and a short characterizing portion (thus a short claim) is the most ideal situation, and that long preambles with long characterizing portions (thus forming long claims) are the least ideal.

Attorneys had more difficulty distinguishing the other two situations: short preamble & long characterizing portion versus long preamble & short characterizing portion. The former situation seemed unrealistic. A short preamble would signal a relatively new technology, while the lengthy characterizing portion would signal only an incremental improvement. Basically, it means the applicant failed to write a good patent claim, since it was likely that the scope could have been defined wider. More interesting analyses were possible on the second situation. A large preamble could signal a more mature technology field. This could be in favor of value, since it normally applies to a larger market. However, this also increases the likelihood that the claim is only describing an incremental improvement. Hence, the size of the characterizing portion is importance. A shorter characterizing portion could signal a wider scope, which would be positively related to value.

One attorney drew a picture (shown on the next page) explaining the difference between these two situations. The size of the white ovals represents an 'unclaimed' space left in a certain field of technology in which the patent places itself. The size of the checkered little ovals represents the scope or claimed area of that patent in the field. For long preambles, little unclaimed space is expected, but with a short characterizing portion the claim will be able to claim a larger part of it. In the other situation, more unclaimed space is expected but due to the long characterizing portion the patent seems less able to claim it.

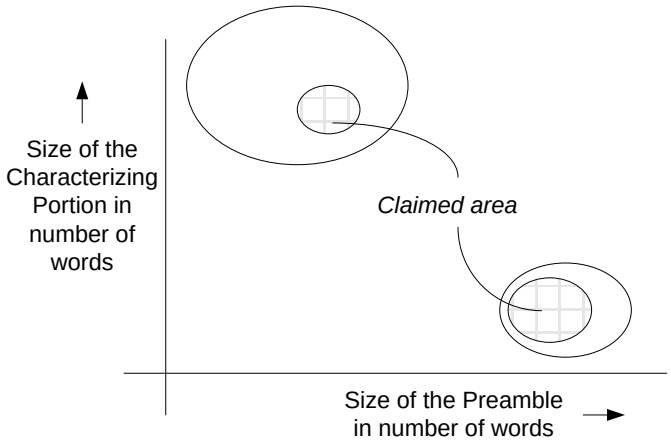

Figure 11-1 - Relation between length of preamble and characterizing portion

Since there was great debate about the exact relation between the structure of the preamble – characterizing portion and patent value, the following hypothesis is put forward:

- **There is a relation between the structure of the Preamble and the Characterizing Portion in the claims of the claim set of a patent and the value of that patent.**

11.2.7 The Number of Limiting Words

The style of writing used in patents differs greatly from the common writing style. Formulated in a legal fashion, patent claims have their own jargon, as the example on page 19 shows. Some particular 'patent' jargon was believed to be related to the introduction of additional restrictions in a claim. These are phrases as 'comprising', 'comprises', 'wherein', 'whereby', 'in which', 'constituted by' and 'consisting of'. They introduce certain features into a claim and therefore narrow its scope. It has been suggested that the amount of these 'limiting words' could indicate the wideness of the scope and therefore have a negative relation to value. Of course, one has to correct for the total number of words in a claim using this variable. The following hypothesis was put forward:

- **There is a negative relation between the amount of limiting words in the claim set of a patent and the value of that patent**

11.2.8 Applications and Grants

In scientific literature there is little attention for the difference between patent applications and grants. For instance, Reitzig (2004a) mentions that claims are attractive as indicator since they are available early in time, namely directly after publication of 'the patent'. However, it is the application that is published. This means that the claim set can still change in its examination process. An exception to this is the article of Guellec et al. (2000), who examines the likelihood of applications being granted in the EPO. However, the possible differences between applications and grants are not discussed there. However, patent attorneys stressed the importance of acknowledging these differences. They strongly prefer to look at grants. One attorney event went so far as stating that looking at applications to predict value is 'lunacy'. Therefore, the following hypothesis was put forward. Note that if this hypothesis cannot be rejected, it means that early prediction with patent claims becomes problematic.

- **There is hardly any relation between the claim set of a patent application and the value of the granted version of that patent**

Attorneys regard the applications as a sort of 'opening bluff' in which the applicant tries to claim as broad as possible, after which the examiner will restrict the scope of the claim to certain proportions if this is needed. Therefore, the amount of differences could be expected to have a negative relation with patent value: since changes can indicate that the applicant did not succeed in obtaining the patent in the format it was desired, the value of the patent fell. A second argument would be that examiners try to narrow down the scope of a claim. So differences could signal that during the examination process, the more narrow scope was approved, which could signal diminished value.

On the other hand, some patent attorneys mentioned that the number of differences could indicate a higher patent value. If changes were made, this would indicate that the scope of the patent was narrowed down to a size which was just acceptable for the examiner. If no changes were made, it could thus mean that the claimed scope theoretically could have been wider. Thus, changed patents could indicate a maximized scope and thus, on average, higher patent value. However, not all attorneys agreed on this line of reasoning.

A second argument for this alternative hypothesis suggests that patents that are expected to be of high value are filed with a large amount of claims and as broad as possible. One could expect that for patents with a high expected value, applicants would take more risk in filing a scope as broad as possible since it is probably worth the extra effort of reformulation and dropping claims. This sort of filing behavior makes changes more likely. Thus, there should be a relation between patents of high expected value and differences between applications and grants. For patents with a low expected value, applicants could prefer to play safe and file in such a way that changes can be kept to a minimum. Thus, this would lead to fewer changes for patents with lower expected value. Following this line of reasoning, it is thus not expected that the loss of claims will lower the value of the patent much. A third and last argument starts with the assumption that higher patent value can be expected in active, more mature markets. However, this will also cause more prior art to be present and thus the likelihood that examiners will not agree with claims rises. Thus, this will cause more differences.

Thus, two contradicting hypothesis could be formulated. If the size of the scope is believed to be related with patent value, more differences could signal lower patent value. Therefore, the following hypothesis was put forward:

- **There is a negative relation between the amount of differences between patent applications and grants and the value of that patent**

However, if it is expected that patents with a higher expected value will probably face more changes during the examination process, than the following hypothesis is put forward:

- **There is a positive relation between the amount of differences between patent applications and grants and the value of that patent**

11.3 Expert Opinions on Statistical Patent Valuation

Beforehand, it was expected that patent attorneys would be skeptical about the use of statistical patent valuation. Their work consists mainly of detailed analyses on individual patents. However, it turned out that the interviewed attorneys were accepting the idea of statistical patent valuation. They, too, are aware that the changing patent world requires new analysis tools in order to keep up with its challenges like the increasing number of patent applications. According to one of the interviewees, this sort of *'intelligent gambling'* can help. To what degree statistical patent valuation can assist attorneys in their work is a different story. Using statistics to select a part of a portfolio or to rank a portfolio, so that attorneys can go through large sets more efficiently, is seen as the main

advantage that can be offered at the moment. However, the systems should never be used as base for decision making on to make statements about individual patents.

Knowing this, it was a bit surprising to find out that several of the interviewees had not worked with PLATO output yet. Some others have used the system's output. In one interview, an IP Counsel expressed how Overall Score was used to make a selection in a very large portfolio that needed to be analyzed. Although sometimes the system ranked important patents too low, he did not encounter any crucial mistakes by the system. Another positive aspect was that for this project, PLATO greatly reduced the workload. For comparison, consider the story of another attorney that explained how such a project was approached at her former employee. Here, a set of say 20.000 patents has been analyzed. First one attorney made a pre-selection merely by looking at titles, after which the remaining about 1000 patents were again considered for relevance by looking at their abstracts. This resulted in around 300 patents which were analyzed by their claims, leaving 50 to a 100 patents for a final in-depth analysis. It can be questioned to what extent human performance in the first two steps would surpass a statistical analysis. The benefit of a better (human) discriminator could easily be cancelled out by the human errors made. Thus, the statistical tools have the advantage that they can evaluate much more (diverse) data to come to the same selections, especially in the first step.

One major argument against statistical patent valuation mentioned by several attorneys was that indicators fail to include any sort of market information. Patent value is to a large extent determined by market conditions. For example, it was mentioned that within consumer electronics there is a relatively low entry barrier and a lack of unique products. Here, having patents in a standard that is commonly used is more important than having strong blocking positions. However, statistics fail to apply for these kinds of criteria. One IP Portfolio Manager further expressed his concern for giving individual patents a score. Those creating scores have a certain intentional use for them. However, they might differ with the ones actually using the scores. There is always the risk that under time pressure the scores are used for more than they are intended to signal.

According to the interviewees, statistical patent valuation (thus systems like PLATO) could be used in the following business applications within Philips:

A) Mergers and Acquisition of other companies – if a business sector of Philips identifies a certain company that is of interest to them, their IP portfolio could be quickly analyzed using statistical patent valuation.
B) Benchmarking – statistical patent valuation is a great tool to benchmark one's IP portfolio with that of competitors. For comparison: in the past this was done by merely counting the number of rights.
C) Identification of white spots or highlighting – within portfolios, highlights can be made by to identify the most interesting patents. Also white spots can be identified, which are 'areas' in the patent landscape of a technology field that are not protected by patents yet.
D) Scouting process – when looking for patents for sale and / or assessing them, statistical patent valuation may provide a first selection criteria or an extra insight to make the evaluation.

11.4 The 'Reitzig article' discussed

This final part presents the view of some attorneys on five statements made in the Reitzig article of 2004 named '*Improving patent valuations for management purposes – validating new indicators by analyzing application rationales'.* These statements were discussed for two reasons. First, they helped to start discussions during the interviews. Second, the statements relate to the latent determinants which are believed to be operationalizable by claims according to Reitzig. Thus, they could be relevant for this thesis project.

In his article, Reitzig gives five statements on page 944 and 945, presented here below. These were shown to some attorneys during the interviews. In general attorneys accepted the statements as being true in some situations but not in all situations.

A) Little state of the art hints at maximum at a 'latent' market where benefits from patenting can be expected in the future.

Attorneys regarded this as a 'statement of fact' with no further remarks.

B) Comprehensive state of the art points at an active market and patenting seems profitable. However, an increasing state of the art raises the risk of legal conflict with competitors and therefore decreases the expected profits.

Attorneys agreed also with this statement. The difficulty in the presented market situation here is to actually find something that you can patent. Also, the scope of claims will become narrow if there is increasing state of the art, which makes enforceability problematic. This is because there are probably plenty of alternatives for competitors so that monopolizing a specific technology 'field' is difficult.

C) If inventive step is small and there is little state of the art, expected profits are small.

Ceteris paribus, this statement may hold. When this assumption is dropped, the size of the market that the invention will serve is mostly determining the profit rate. Since little state of the art relates to a new technological field with a yet to be developed market, the expected profits on the short run may indeed be small. However, expected profits for the future might be high. However, in general it is not expect that under the conditions of small inventive step and little state of the art, patents are filed. If there is little state of the art, considerable inventive steps can be expected to follow soon. It is expected that applicants will wait for them before filing applications.

D) If inventive step is small and there is comprehensive state of the art, possible profits are high. However, the risk of losing the patent in a legal argument rises, too, decreasing the overall expected profits from patenting. Expected profits may range from medium to high.

If there was an extensive and thorough search for prior art, but the inventive step has survived these searches, it indicates that the patent will have a strong position in case of litigation. Combined with comprehensive state of the art (indicating market opportunity) indeed profits can be high.

E) If inventive step is high and there is comprehensive state of the art, possible profits seem high, and there is little risk of losing the patent in a legal argument. Expected profits are very high.

This is considered a valid statement as long as the invention is desired by the market. Therefore, it should be an invention with few alternatives making it difficult for competitors to 'invent' their way around the patent.

12 Construction of Datasets

The interviews delivered a set of the claim characteristics that could possibly indicate patent value. From here onwards, the paper will focus on providing statistical evidence for these presumed relations. For this, quantitative data had to be gathered on both claim characteristics and patent value. Once such a dataset is constructed, statistical methods can be used to look for significant relations.

In this chapter, the construction and content of two datasets is described. The first set contains all data to test the relation between claim characteristics and patent value. The second set was constructed to determine the effect of external influences, which was difficult to do with the PatVal set as we will see.

12.1 Construction of the PatVal Set

To test the relation between claim characteristics and patent value, a dataset should contain at least a set of quantitative claim characteristics of each patent and a value indication of these patents. However, since literature has identified that multiple dimensions of patent value exist, the use of several value indicators is preferred.[7] Fortunately, this project was offered access to a set of value estimations of patents. In a research project within six European countries, almost 30,000 inventors that produced an EP patent between 1993 and 1998 were asked the following question about that patent: *"Suppose that on the day in which this patent was granted, the applicant had all the information about the value of the patent that is available today. In case a potential competitor of the applicant was interested in buying the patent, what would be the minimum price the applicant should demand?"* (Giuri et al., 2007, p. 1120). This delivered around 9,000 value estimations making this set unique by its sheer size. The research project was described in an article named 'Inventors and invention processes in Europe: Results from the PatVal-EU survey'. Therefore, this dataset is called the 'PatVal' set. An overview which patents are related to the business areas of Philips can be found in table 12-1 and is based on IPC-classes.

IPC Class	Content
A61	Human Necessities – Health – Medical or Veterinary science, Hygiene
C09	Chemistry – Dyes, paints, polishes, natural resins, adhesives, compositions not otherwise provided for, applications of materials not otherwise provided for
F21	Mechanical engineering, lighting, heating, weapons, blasting - Lighting
G01	Physics – Instruments – Measuring, Testing
G02	Physics – Instruments – Optics
G03	Physics – Instruments – Photography, cinematography, analogous techniques using waves other than optical waves, electrography, holography
G06	Physics – Instruments – Computing, Calculating, Counting
G11	Physics – Instruments – Educating, Cryptography, Display, Advertising, Seals
H01	Electricity – Basic Electric Elements
H02	Electricity – Generation, conversion, or distribution of electric power
H03	Electricity – Basic Electronic Circuitry
H04	Electricity – Electric Communication Technique
H05	Electricity – Electric techniques not otherwise provided for

Table 12-1 Overview of IPC-classes of PatVal set

This subset has already been used within IP&S to examine the relationship between these value estimations and the Overall Score. This subset contained value estimations on 2703 granted patents in technology fields relevant to

[7] This is due to the findings in literature that patent value has several dimensions. See for a report on this topic: Van Zeebroeck, 2008 – 'The puzzle of patent indicators'

Philips. These are (1) Electrical Engineering with 1462 patents, (2) Instruments with 846 patents and (3) Chemicals and Pharmaceuticals with 393 patents.[8] Since there was no rationale for using any other set, this subset was also used for this project.

Next to the empirical value estimate from the PatVal project, patent value indicators based upon 'first-page' information were used. Including such data makes it possible to assess the contribution of a claims indicator compared to the traditional, already existing indicators. In addition, the relation of a claims indicator with the traditional indicators can be examined. It will be interesting to see with which traditional indicators the claims indicator will have most in common, especially since the use of a single value estimation is considered problematic (see Zeebroeck, 2008). This argument, primarily aimed at the use of patent indicators as value estimators, claimed that different indicators measure different dimensions of value. If claims can predict patent value but have a relatively low relation with the traditional indicators, this can indicate that they cover other dimensions of value than the traditional indicators do. A fairly independent indicator would be a welcome contribution to the arsenal of patent indicators.

To construct traditional patent indicators, family size and opposition data were extracted from MicroPatent. For other traditional indicators, their indicator in PLATO was used. The most important difference between using raw data and the PLATO indicators is that PLATO uses non-linear transformations to create indicators with a range from zero to ten. These transformations are not emperically validated and are therefore, to some extent, questionable. However, the exact methods of transformation cannot be given here due to confidentiality reasons. The Forward Citations Indicator in PLATO uses the number of forward citations received by a patent family and corrects for the age of the patent. The Backward Citations Indicator and the Status Indicator were also selected so that all major PLATO indicators were included in the dataset. The Backward Citation Indicator is based upon the number of backward citations but weights them using X- or Y- citations, if available. To obtain the Status Indicator, first the legal status of each family member is given a certain value. The Status Indicator is than the average value of all members. Although the PLATO indicators are not identical with the traditional indicators, they will be treated as if they were. Finally, the Overall Score, which is a composite index of several of the PLATO indicators, is included. It will be used on several occasions next to the value estimations of the PatVal survey to see to what extent the Overall Score is related with claim characteristics.

12.1.1 Quantifying Claim Characteristics

A second part of the dataset construction consisted of quantifying the claim characteristics. Many characteristics were identified like the number of claims, the length of claims and the difference between independent or dependent claims. Also claims could be product or method claims, contain a certain amount of limiting words and be formulated more functional or more in a 'portrait' fashion. Finally, the two-part form structure could hint in some cases towards patent value.[9]

For all PatVal patents, the claim set was extracted from the database MicroPatent. This database delivered the claims in an Excel format. In Excel, scripts were written that could process these claims. These scripts delivered per claim a quantitative value for each claim characteristic. The following table provides an overview of the way how

[8] Of two patents, the data on technology field was missing

[9] Patent attorneys also stressed the difference between applications and grants. For now, only grants will be taken into account. Value predictions based upon applications or the differences between applications and grants will be discussed in chapter 15

different claim characteristics were quantified. Note that some scripts make use of the English language, which makes them inappropriate to use on claims written in another language without translations.

Claim Characteristics	Method of Calculation
Number of claims	Counting of all filled cells
Number of words in a claim	Counting the number of characteristics minus the number of spaces
Independent vs dependent claim	In case a reference was found in a claim, it was identified as dependent. If not, the claim was seen as independent.
First Claim	Looking only at the first claim of the claim set
Grant versus Application	These were separated in the dataset itself
Product claim	In case the words 'apparatus', 'system', 'device', 'circuit', 'source', 'unit', 'amplifier' or 'detector' were used in the first part of an independent claim, it was considered a product claim.
Method claim	In case the words 'method' or 'process' were used in the first part of an independent claim, it was considered a method claim.
Functional claim	In case the phrase "for ….ing" was used in an independent claim, it was considered a functional claim.
Functional phrases	Counting the amount of phrases "for ….ing" that were used.
Limiting words	Counting the amount of times that one of the following words was used: 'comprising', 'comprises', 'wherein', 'whereby', 'in which', 'constituted by' and 'consisting of'.
Preamble and Characterizing Portion	Counting from the start, the first time when either the word 'characterizing' or 'characterising' was used (which differ in the use of a 'z'or an 's'), it identified the end point of the preamble and the start of the characterizing portion.

Table 12-2 Overview of rationales of behind the scripts quantifying claim characteristics

Some claim characteristics cannot be transformed into quantitative values with 100% accuracy. This was due to the fact that information contained in text needs a high level of sophistication for successful transformation. For instance, to identify the number of product claims a script looked for the use of specific words in the first part of an independent claim. However, every independent claim using a noun in its first part could be a product claim. Since no methods of identifying nouns in text are available, the risk of some inaccuracy had to be accepted.

Finally, some characteristics had an additional problem related to the level of analysis. Analysis can be done on the claim level or on the patent level. Since value estimations were only available on the patent level, characteristics that were available on the claim level had to be transformed to the patent level. Most of the time, this was done by averaging, using the median or using summation.

12.2 Construction of the Philips-Electronics dataset

A second dataset was constructed to examine the influence of external factors on claim sets as for instance country or technology differences. For this set, no value estimations from PatVal were available and therefore, the set could only be used to test these factors. In total, six external factors were identified by literature and the interviews. Of them, the following three were not taken into account in this thesis project:

A) Language differences – this thesis project focuses only on patent claims that are written in or translated into English.[10]

B) Fee height for additional claims – no radical changes were found in the fee policy of the EPO (and the USPTO) with respect to fees on additional claims until 2008. In 2008, fees on additional claims were highly increased in

[10] The only exception on this rule was made when examining applications on basic characteristics as number of claims or number of words. This will be explained in further detail in chapter 15 dealing with value predictions with the use of applications.

the EPO and it is expected to happen soon in the USPTO. Therefore, research including the year 2008 or later years might have to take into account this policy change. Patent attorneys mentioned in interviews that the policy on writing claims within Philips was indeed altered as a result of the policy change of the EPO.

C) The applicant's writing style and expertise – Writing a patent is like painting. Give two artists the assignment of making a portrait of the same person and the paintings will never be 100% equal. The exact formulation and looks of a patent is dependent on the patent attorney's style of writing and his or her expertise. Unfortunately, due to the fact that data on such issues was not available, this was not taken into account.

Three other factors were considered to be of major importance and were reasonably feasible to examine. These three factors relate to differences in claim sets caused by the country the patent originates from, the technology field it describes or the year in which it is filed:

D) Country differences – it has been assumed that there is a difference in the structure of a claim set between countries. This is most likely caused by the fact that patent law differs across countries, accepting different numbers of claims or prescribing a certain structure in which the patent should be written.

E) Technology field differences – several patent attorneys confirmed literature findings in the fact that claim sets stemming from different technology fields have a different 'look'.

F) Time differences – although it was expected that claim sets have altered throughout time in some ways, it was not clear what kind of changes have occurred over time.

The PatVal set fails to control for country differences as it focuses only on EP patents. Second, time differences will also be marginal since all patents were filed between 1993 and 1997. Finally, the Philips-related subset of PatVal was unequally spread across the three technology fields it contained. Therefore, a second dataset was constructed.

This set contains 3000 granted patents which were gathered in such a way that possible country, technology and time differences could be measured. First, three technology fields were identified matching three business fields of Philips, hence explaining the name of the dataset: 'Philips-Electronics'. Second, in each field 750 patents were selected over a time period of 20 years. Of them, 250 had at least a US family member but no EP family member. 250 others had at least an EP family member but no US family member, and the last 250 had both an EP and a US family member.

To clarify, the following table is sketched:

	US but no EP	**EP but no US**	**US and EP**	*Sum*
Optical Storage (Consumer Lifestyle)	250 US grants	250 EP grants	250 US grants and 250 EP grants	1000 grants
Medical Imaging (Healthcare)	250 US grants	250 EP grants	250 US grants and 250 EP grants	1000 grants
Lighting Electronics (Lighting)	250 US grants	250 EP grants	250 US grants and 250 EP grants	1000 grants
Sum	750 grants	750 grants	750 grants	3000 grants

Table 12-3 Overview of the content of the Philips-Electronics dataset

The gathering of claim characteristics and traditional patent indicators for the Philips-Electronics set happened in the same way as was described for the PatVal set.

13 Analysis - Exploration of the dataset and the influence of external factors

This chapter serves as an exploratory preparation for the actual statistical analyses testing the relation between claim characteristics and patent value. This preparation is threefold. First, the quantitative values of the claim characteristics were explored. Second, the values for the value estimations and traditional patent indicators were presented. Finally, the influence of the external factors (country, technology, time) was tested.

13.1 Exploration of the claim characteristics

This part gives the reader an understanding of the values of the quantified claim characteristics. We start with some basic variables. Table 13-1 shows that 90% of all patents have a claim set in the range of 1 to 20 claims. Interestingly, 20 is also the number of claims that can be filed in the US without additional payments. (for the EPO, that number was 10 but has recently changed to 15) Only 2.5% of all patents contains over 30 claims. The number of independent claims is low. Just 5% of all patents have three or more independent claims.

Number of Claims			**Number of Independent Claims**			**Number of Dependent Claims**		
	Frequency N = 2703	% Frequency		Frequency N = 2701	% Frequency		Frequency N = 2701	% Frequency
1 - 5	524.00	19.39%	1	1846.00	68.35%	0	37.00	1.37%
6 - 10	1110.00	41.07%	2	654.00	24.21%	1 - 5	750.00	27.75%
11 - 15	581.00	21.49%	3	116.00	4.29%	6 - 10	1060.00	39.22%
16 - 20	267.00	9.88%	4	55.00	2.04%	11 - 15	479.00	17.72%
21 - 25	119.00	4.40%	>5	30.00	1.11%	16 - 20	204.00	7.55%
26 - 50	93.00	3.44%				21 - 25	94.00	3.48%
>50	9.00	0.33%				26 - 50	70.00	2.59%
							>50	0.33%
Mean	10.93		Mean	1.45		Mean	9.48	
Median	9.00		Median	1.00		Median	8.00	
Max	92.00		Max	19.00		Max	91.00	

Table 13-1 Overview of the content of the PatVal set – number of claims

The total number of words seems normally distributed but skewed to the left: abound 70% of the patents is within the range of 200 to 800 words, while there is is a minority with outliers up to almost 5000 words[11] (see table 13-2 on the next page). Independent claims are longer than average claims. Claim lenghts for independent claims of over 200 words are no exception and outliers up to 800 words in a single claim are recorded. This means that they are much longer as the dependent claims. (the average length of a dependent claim was 42 words)

[11] The number of claims and the number of words total are .780 correlated at a significance level of .01 (2-tailed).

Number of Words Total			**Number of Words Average**			**Number of Words Average in Independent Claims**		
	Frequency N = 2703	% Frequency		Frequency N = 2701	% Frequency		Frequency N = 2701	% Frequency
1-200	119.00	4.40%	1 - 25	33.00	1.22%	0-50	159.00	5.89%
201-400	706.00	26.12%	26-50	960.00	35.52%	51-100	575.00	21.29%
401-600	708.00	26.19%	51-75	1004.00	37.14%	101-150	693.00	25.66%
601-800	511.00	18.90%	76-100	389.00	14.39%	151-200	602.00	22.29%
801-1000	294.00	10.88%	101-125	178.00	6.59%	201-250	356.00	13.18%
1001-1500	244.00	9.03%	126-150	57.00	2.11%	251-300	157.00	5.81%
>1500	121.00	4.48%	151-200	44.00	1.63%	301-400	110.00	4.07%
			>200	38.00	1.41%	401-500	32.00	1.18%
						>501	17.00	0.63%
Mean	646.46		Mean	67.11		Mean	157.71	
Median	547.00		Median	57.90		Median	145.00	
Max	4940.00		Max	705.00		Max	839.00	

Table 13-2 Overview of the PatVal-set - number of words

The other claim characteristics will be explored into less detail in table 13-3. A few interesting facts can be seen. There is a larger number of product claims than of method claims. This could be true, but also be caused by the fact that the script worked better for product claims. The mean for number of words in the first claim is almost equal with that for independent claims. This tells us that the first claim does not differ much from the average independent claim. The preamble and the characterizing portion have almost the same length, although the characterizing portion seems to have higher outliers.

	Number of Product Claims[12] N = 2703	**Number of Method Claims** N = 2703	**Number of Functional Claims** N = 2703	**Number of Functional Phrases in Indepen-dent Claims** N=2701	**Number of Limiting Words in Indepen-dent Claims** N = 2701	**Number of Words in First Claim** N = 2703	**Number of Words in the Preamble** N = 2006	**Number of Words in the Characteri-zing Portion** N = 2006
Mean	0.53	0.38	0.76	1.25	5.12	160.40	83.42	84.47
Median	0.00	0.00	1.00	1.00	4.00	147.00	72.58	71.00
Std Dev	0.72	0.61	0.83	1.73	3.96	91.09	57.14	59.13
Min	0.00	0.00	0.00	0.00	0.00	10.00	1.00	9.00
Max	10.00	6.00	14.00	20.00	25.00	897.00	377.00	627.00

Table 13-3 Overview of the PatVal-set - other characteristics

[12] In several cases, claims were identified as being both product and method claims, although they can be only one of the two. Unfortunately, this error was identified after all the data analysis was conducted and thus the erroneous cases were not excluded.

Finally, some ratios were calculated based upon the claim characteristics. They can provide additional insights on the relation with value. Their values are sketched in table 13-4 shown below. Interestingly, the ratio of independent claims versus all claims differs between measuring it with words or claims. For words, the average ratio is 0.38 while for claims it is 0.18. The ratios for limiting words and functional phrases are very low, even when looking at their maximum scores.

	Ratio # of words in Independent claims vs total # of words in all claims N = 2703	**Ratio # of words in Preamble of First Claim vs Words in First Claim** N = 2703	**Ratio # of Independent claims vs total # of claims** N = 2703	**Ratio of average # of Words in Independent claims vs Dependent claims** N = 2664	**Ratio of # of average limiting words vs average # of words (in Independent claims)** N = 2701	**Ratio of # of average functional phrases vs average # of words (in Independent claims)** N = 2664
Mean	0.38	0.36	0.18	3.62	0.03	0.01
Median	0.34	0.39	0.14	3.24	0.03	0.01
Std Dev	0.20	0.28	0.15	1.98	0.02	0.01
Min	0.00	0.00	0.00	0.14	0.00	0.00
Max	1.00	0.96	1.00	19.47	0.18	0.07

Table 13-4 Overview of the PatVal-set - ratios

13.2 Exploration of the value estimations and traditional patent indicators

An overview is presented of the value value estimations and traditional patent indicators. The PatVal value clearly shows the skewed value distribution that one would expect. The same distribution can be found for Family Size. However, this skewness is not visible in the Overall Score, which approaches a normal distribution with high kurtosis. Opposition cases are only incidental as 93% of all patents did not face opposition.

	PatVal Value N = 2275	**Overall Score** N = 2703	**Status Indicator** N = 2703	**Opposition** N = 2703	**Forward Citation Indicator** N = 2703	**Backward Citation Indicator** N = 2703	**Family Size** N = 2703
0	…	…	0	2502	594	120	…
1	197	0	6	161	192	…	13
2	374	11	71	27	96	…	131
3	464	442	643	3	94	…	281
4	474	549	947	8	169	…	508
5	339	754	730	0	508	50	491
6	196	774	164	0	348	…	373
7	96	163	72	0	432	…	210
8	47	9	63	0	158	…	170
9	18	1	0	0	44	…	129
10	13	0	7	0	68	2532	94
> 11	57	…	…	2	…	…	303

If dots (…) were printed, it meant that this value could theoretically not be in the set

Table 13-5 Overview of the PatVal-set - value estimations and traditional indicators

The number of actual PatVal values in the set is significantly lower than was expected. This is caused by the fact that not every inventor that participated in the PatVal-survey had assigned a value to the patent under investigation. Overall, 13% of the PatVal set lacked a value estimation; here the percentage is almost 16%, which is in the same order of magnitude.

13.3 Examining the influence of external factors

This paragraph determines if external factors influence the content of a claim set. This was done with the use of the Philips-Electronics dataset[13]. First, country differences were examined by comparing US and EP patents belonging to the same patent family. Second, it was checked whether claim sets change over time. Finally, differences due to technology fields are discussed.

The analysis to test for country differences was done by comparing EP and US grants belonging to the same patent family. From table 13-6 it becomes clear that there are substantial differences between US claim sets and EP claim sets. American claim sets have, on average, more independent and dependent claims as their European counterparts. Also, their length in words shorter. To see if these differences are significant, a Wilcoxon Signed Ranks Test was used[14]. A Wilcoxon-Signed Ranks test was conducted to compare the means of EP and US claim sets on all four variables. Significant differences were found. Therefore, it has to be concluded that the claim sets of EP and US patents differ from each other.

						Wilcoxon-Signed Ranks Test	
		Mean N = 743	Min N = 743	Max N = 743	Std Dev N = 743	Z-value	Asymp. Sig. (2-tailed)
Number of Claims	EP	9.32	1.00	64.00	7.52	12.21	.000
	US	13.77	1.00	76.00	11.69		
Number of Independent Claims	EP	1.80	1.00	11.00	1.24	15.07	.000
	US	3.17	1.00	32.00	2.94		
Number of Words Average	EP	67.72	1.14	538.00	50.22	4.34	.000
	US	63.23	1.24	434.67	43.43		
Number of Words average…	EP	147.98	1.00	613.00	102.45	3.50	.000
…in Independent Claims	US	137.74	1.00	583.00	92.64		

Table 13-6 Comparison between EP and US claim sets

Second, the possibility of time dependency is examined. The graphs on the next page show the average number of (independent) claims over time for EP and US patents belonging to the same family. Again, the differences

[13] The absolute values of this dataset related to the independent claims cannot be compared with the PatVal set. This is due to the fact that after the construction of the XXX dataset, the script to identify a claim as independent or dependent was adjusted. The script used for the XXX dataset structurally overestimated the number of independent claims.

[14] Normally, a Paired-Samples T-test would be used. However, a One-Sample Kolmogorov-Smirnov test has rejected the hypothesis that the means for both groups are normally distributed. Since this was one of the requirements for the Paired-Samples T-test, the Wilcoxon Signed Ranks Test was used instead. Also, if in later analyses the Wilcoxon Signed Ranks Test is used, it was because of this reason.

between EP and US patents can be noticed. For the years, the publication dates were used. The values are the average values for all patent published in that period of time.

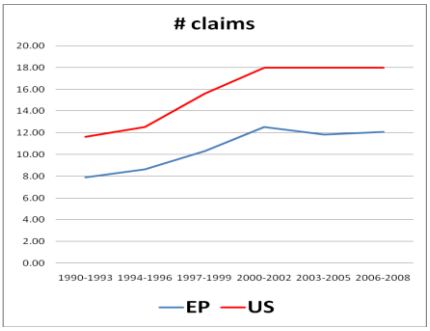

Figure 13-1 – (left) - Number of claims in EP and US patents over time

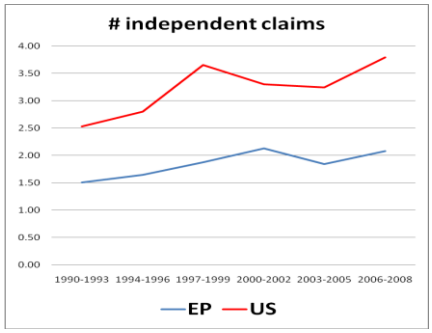

Figure 13-2 - (right) - Number of independent claims in EP and US patents over time

There seems to be an increasing number of claims and independent claims over the years. For further analysis, four time categories were constructed. EP and US patents were analyzed seperately. Table 13-7 presents the means for the number of claims and claim length in in each time category. Also the results of a Kruskal-Wallis test[15] are given, showing that in 7 out of 8 cases differences are significant. Only for the number of words in independent claims in EP patents, the hypothesis that the values are equal in all four time categories could not be rejected. Therefore, it is concluded that claims sets have a time dependency.

		Time Categories				Kruskal-Wallis Test	
		1987-1992 N = 55	1993-1998 N = 264	1998 – 2003 N = 193	2004-2008 N = 231	Chi-Square	Asymp. Sig. (2-tailed)
Number of Claims	EP	7.65	7.92	10.20	10.69	19.53	.000
	US	11.60	11.38	14.70	16.24	26.32	.000
Number of Independent Claims	EP	1.55	1.73	1.79	1.97	20.43	.000
	US	2.71	2.97	3.19	3.48	11.97	.007
Number of Words Average	EP	66.23	74.44	60.44	66.48	8.01	.046
	US	65.12	69.96	60.43	57.44	12.83	.005
Number of Words average…	EP	128.98	157.70	144.76	144.09	4.61	.203
…in Independent Claims	US	146.59	154.72	128.99	123.53	16.86	.001

Table 13-7 Comparison of claim sets stemming from different periods in time

Finally, the assumption that differences between technology fields exist was tested. The Philips-Electronics set contains three technology fields (Lighting Electronics, Medical Imaging, and Optical Storage). A table with their means and standard deviation for all four variables (both for EP and US) is presented on the next page. Structural differences between technology fields seem to exist, although a steady trend cannot be identified. For instance,

[15] Normally, ANOVA would be used. However, a Test of Homogeneity of Variances showed that the assumption that all groups have equal variances was not met. The Kruskal-Wallis test does not need this assumption. Instead, it requires that there is a similar distribution of values in all variables. Boxplots showed this was (fairly) the case.

both for EP and US patents, MI scores higher on the number of claims, while OS scores higher on the number of independent claims. A Kruskal-Wallis test gave significant differences for the number of independent claims and number of words average, but not for the number of claims. The number of words average in independent claims was not significant in the EP case, but was significant for US claim sets. An additional problem is that MI was overrepresented in more recent years and so time dependency could have influenced the findings.

		LE		MI		OS		Kruskal-Wallis Test (1)	
		Mean N = 249	Std Dev	Mean N = 245	Std Dev	Mean N = 249	Std Dev	Chi-Square	Asymp. Sig. (2-tailed)
Number of Claims	EP	9.88	7.19	11.18	7.13	10.80	7.41	4.38	.11
	US	15.02	10.53	16.60	11.97	15.56	11.87	2.65	.27
Number of Independent Claims	EP	1.57	1.40	1.76	1.11	2.23	1.96	33.95	.00
	US	2.84	2.32	3.28	3.44	3.60	3.07	9.50	.01
Number of Words Average	EP	76.83	40.57	72.97	41.39	88.59	56.51	16.25	.00
	US	72.56	42.66	67.44	36.37	84.00	51.18	21.69	.00
Number of Words average…	EP	176.47	77.16	179.38	96.07	185.17	94.38	1.19	.55
…in Independent Claims	US	165.09	68.64	156.65	78.14	172.37	85.96	6.37	.04

Table 13-8 Comparison of claim sets stemming from different fields of technology

What to conclude out of this table? Since the results are not giving any clear picture, it has been decided that there will be a correction for technology field dependency. However, from these results it cannot be concluded that there are significant differences between the technology fields LE, MI and OS.

13.4 Conclusion

The first part of this chapter gave the reader an impression of the content of the PatVal dataset. In particular, it included some lessons on the typical characteristics of claim sets and the claims they contain. Further, the PatVal set was added with three dummy variables. One dummy takes into account time dependency. If the patent was filed between 1996 and 1998, the dummy had value one, while for the time period 1993-1995, a zero was given. For technology fields, two dummies were included since the set holds three technology fields: Electrical Engineering, Instruments and Chemicals&Pharmaceuticals. Since the largest population was Electrical Enginering, this was used as base.

14 Analysis – Constructing a value indicator based upon patent claims

The preceding chapter gave an overview of the values of all variables. Also, it concluded that in further analyses, one has to control for time, technology field and country dependencies. This chapter presents the findings of tests that examined which claim characteristics can be used to predict patent value. In addition, their contribution compared to the traditional indicators was tested.

First, the correlation of the claim characteristics with PatVal Value and Overall Score was tested. Second, by using principal component analysis, the number of characteristics was reduced to several components. These components were used in a series of regressions to test if they can predict patent value. Also, for one regression model, the raw variables underlying one of these components were used. Next, with the use of ordinal logit regression, the marginal effects for the significant value predictors are calculated. Finally, using in one model both claim characteristics and traditional indicators, the contribution of using claim characteristics was examined.

14.1 The correlation of claim characteristics with value

The correlation matrix below shows which claim characteristics are correlated with PatVal value or PLATO's Overall Score. For most variables the sample size is N = 2703, however for some cases less variables were available. Further, the expected relation with patent value is shown which was based upon the hypotheses described in chapter 11.

Correlation Matrix

	PatVal Value	Overall Score	Expected relation
# of Claims	.133 **	.213 **	+
# of Independent Claims	.034	.074 **	+
# of Dependent Claims (1)	.133 **	.210 **	+
# of Product Claims	-.058 **	.010	+
# of Method Claims	-.005	.055 **	+
# of Functional Claims	-.027	.020	+
# of Words Total	.072 **	.144 **	+
# of Words Average	-.068 **	-.084 **	-
# of Words in Independent Claims Total	-.020	.026	-
# of Words in Independent Claims Average (2)	-.051 *	-.022	-
# of Words in Independent Claims Minimum (2)	-.049 *	-.029	-
# of Words in Dependent Claims Total (3)	.093 **	.159 **	+
# of Words in Dependent Claims Average	-.021 **	-.041 *	-
# of Words in the Preamble Average	-.051 *	.014	?
# of Words in the Characterizing Portion Average	-.003	.031	-
# of Words in Product Claims Average	.025	.015	-
# of Words in Method Claims Average	-.034	-.031	-
# of Words in Functional Claims Average	-.039	.001	-
# of Words in the First Claim (4)	-.044 *	-.017	-
# of Words in the Preamble of the First Claim (5)	-.043	.026	?
# of Words in the Characterizing Portion of the First Claim	-.033	-.042 *	-
# of Limiting Words Total	.069 **	.140 **	-

# of Limiting Words Average	-.021	-.007	-
# of Limiting Words in Independent Claims Total (6)			-
# of Limiting Words in Independent Claims Average	-.028	-.001	-
# of Functional Phrases Total	.014	.079 **	+
# of Functional Phrases Average	-.037	.008	+
# of Functional Phrases in Independent Claims Total	-.018	.047 *	+
# of Functional Phrases in Independent Claims Average	-.032	.030	+
Ratio of # of Words Total of Independent Claims / All Claims	-.095 **	-.119 **	-
Ratio of # of Words in the Preamble of the First Claim / First Claim	-.057 *	-.008	?
Ratio of # of Independent Claims / All Claims (7) (8)	-.047 *	-.127 **	?
Ratio of # of Words Total Independent Claims / Dependent Claims (8)	-.030	.025	?
Ratio of # of Limiting Words Average / # of Words Average	.026	.034	-
Ratio of # of Functional Phrases Average / # of Words Average	-.008	.017	+

** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

(1) Correlated .993** with the # of Claims

(2) Correlated .976** with each other

(3) Correlated .935** with the # of Words Total

(4) Correlated .980** with the # of Words in Independent Claims Average

(5) Correlated .986** with the # of Words in the Preamble Average

(6) Wrong values due to error in script

(7) Significance depends solely on the # of Claims

(8) The expected relation might be dependent on the size of the claim set. Therefore, it was difficult to assess the relation with patent value

Table 14-1 Correlation matrix of claim characteristics with PatVal value and Overall Score

In general, it seems that some of the more 'sophisticated' claim characteristics do not relate to patent value. The number of functional phrases, as well as the number of limiting words, did not give an indication of patent value. Also the number of functional claims as well as the length of product, method and functional claims was not related to patent value. The length of the preamble and the characterizing portion also showed little significant correlation with patent value.

Rather basic claim characteristics functioned better. As expected, the number of claims was positively related to value. This holds both for the number of total claims as for the number of independent claims. Also the length of claims gave significant findings: it seems to be negatively related to patent value. This indicates that there is some truth in the rule of fist of patent attorneys that longer claims are poorer claims. Finally, the length of the preamble seemed to have a relation with patent value. Interestingly, the ratio between the preamble-length in a first claim and the total length of a first claim was negatively related to value. As a test, a constant was used in the nominator which gave no significant relation. Therefore the effect should be contributed to the length of the preamble. The characterizing portion only related significantly to the Overall Score and is not used in further analyses. The ratio between the preamble and the characterizing portion (not presented here) also gave no significant findings.

Before we proceed, some variables were rejected in further analysis because it seemed their correlation was biased. For instance, the total number of functional phrases (in independent claims) was correlated with the Overall Score. But since there was no correlation with the averages of these variables, it seemed that the relation for the totals was only obtained by their relation with the total size of the claim set. Indeed, they turned out to be significantly related with the total number of words and the total number of words in independent claims, respectively with .484 and .536, both at a 0.01 significance level (2-tailed). The same logic holds for the total number of limiting words. Here, again the average values are not significant, while the total is highly correlated

with the number of words total: .668 at the 0.01 significance level. Therefore, these variables were not used in further analyses.

14.2 Principal Component Analysis

The claim characteristics that gave a significant correlation were used in a principal component analysis, since it is believed that several variables are strongly interrelated. Principal component analysis reduces the number of variables by calculating a component which can represent such interrelated variables with minimum loss of variance. With use of Varimax rotation, the following table was constructed.

Rotated Component Matrix[a]

	Component				
	1	2	3	4	5
# of Claims	-.047	.954	-.016	.139	-.187
# of Independent Claims	-.095	.104	-.045	.859	-.051
# of Dependent Claims	-.037	.962	-.011	.041	-.185
# of Product Claims	-.016	-.006	.257	.544	.147
# of Method Claims	-.027	.107	-.163	.727	.005
# of Words Total	.301	.870	.011	.256	.219
# of Words Average	.596	-.189	.013	.164	.704
# of Words in Independent Claims Average	.973	.047	.085	-.038	.148
# of Words in Independent Claims Minimum	.950	.025	.076	-.150	.131
# of Words in Dependent Claims Total	.098	.951	-.004	.042	.212
# of Words in Dependent Claims Average	.170	.115	-.008	.008	.959
# of Words in the Preamble Average	.677	.011	.705	-.032	.051
# of Words in the First Claim (4)	.957	.055	.090	-.013	.157
# of Words in the Preamble of the First Claim	.666	.022	.719	-.009	.059
Ratio of # of Words Total of Independent Claims / All Claims	.448	-.677	.031	.402	-.145
Ratio of # of Words in the Preamble of the First Claim / First Claim	.006	-.033	.957	.000	-.049

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 6 iterations.

	Initial Eigenvalues		
Component	Total	% of Variance	Cumulative %
1	5.224	32.649	32.649
2	4.084	25.524	58.173
3	1.889	11.807	69.980
4	1.706	10.660	80.641
5	1.106	6.914	87.554

Extraction Method: Principal Component Analysis.

Table 14-2 Principal Component Analysis outcomes of claim characteristics

The principal component analysis delivers five components, together explaining almost 88% of the total variance. The first component seems related to the length of the independent claims (the first claim is always an independent claim). The second component relates to the size of the claim set, measured either in words or in

claims. Also negatively related to this component seems the ratio of the number of words in independent claims compared to the total number of words. Together, these two components already explain almost 60% of the variance in the data.

The third component deals with the length of the preamble. It includes the average length of the preamble and its length in the first claim. Also the ratio between the length of the preamble compared to the total length of the first claim is loading on this component. The fourth components relates solely to the number of independent claims in a claim set. Finally, the fifth component relates to the average length of claims.

14.3 Ordinal Logit Regression

The last two paragraphs have identified five components that are likely to predict patent value. These components were tested on their contribution to predict patent value. In addition, it was examined if they still contribute in a model in which the traditional indicators are also included.

For this analysis, ordinal logit regression was used since the dependent variables (PatVal value and Overall Score) are ordinal variables. There was chosen for a logit regression, since it most fitted the characteristics of the PatVal value. Ordinal logit regression shows if there are significant estimates. Also the sign can be interpreted in the normal way, but unfortunately the size of the estimate cannot be used directly to interpret its effect.

The analysis took place in three steps. First, the traditional indicators were used in a model to predict PatVal Value. Second, the five components underlying the claim characteristics were used to predict both PatVal Value and the Overall Score. Finally, both the traditional indicators and the five components were used in two ordinal logit regressions to predict PatVal Value. For this analysis, the Overall Score was not be used, since it is a composite index of several of the traditional indicators but not of the claims characteristics. This would result in unrealistically high estimates from the traditional indicators.

Ordinal Logit Regression

Location	*Model I*	*Model II*	*Model III*	*Model IV*	*Model V*
Forward Citations Indicator	0.09 (.01) **			0.05 (.01) **	0.08 (.02) **
Backward Citations Indicator	-0.02 (.02)			0.00 (.01)	
Family Size	0.09 (.01) **			0.05 (.01) **	0.09 (.01) **
Opposition	0.24 (.08) **			0.15 (.06) **	0.28 (.11) **
Status Indicator	0.06 (.03)			0.03 (.02)	
Component One		-0.06 (.05)	0.03 (.04)	-0.04 (.03)	
Component Two		0.17 (.04) **	0.34 (.04) **	0.04 (.03)	0.09 (.04) *
Component Three		-0.07 (.04)	-0.03 (.04)	-0.04 (.03)	
Component Four		0.00 (.04)	0.14 (.04) **	-0.03 (.03)	
Component Five		-0.05 (.04)	-0.12 (.04) **	-0.02 (.03)	
Priority Year	-0.03 (.08)	-0.13 (.09)	0.27 (.08) **	-0.06 (.05)	-0.09 (.09)
Technology Field (Instr)	-0.07 (.08)	-0.21 (.10)	-0.01 (.09)	-0.09 (.06)	-0.17 (.09)
Technology Field (ChPha)	-0.77 (.11) *	-0.41 (.18)	-0.74 (.17) **	-0.03 (.11)	-0.12 (.17)
Value					
Value = 1	-1.69 (.28) **	-2.88 (.22)		-0.82 (.21) **	-1.75 (.25) **
Value = 2	-0.40 (.28)	-1.61 (.21)	-5.99 (.38) **	-0.10 (.21)	-0.46 (.24) *
Value = 3	0.56 (.28) *	-0.68 (.21)	-2.07 (.20) **	0.49 (.21) *	0.50 (.24) *
Value = 4	1.47 (.28) **	0.18 (.20)	-0.98 (.19) **	1.04 (.21) **	1.39 (.24) **
Value = 5	2.31 (.28) **	0.93 (.21) **	0.24 (.19)	1.50 (.21) **	2.17 (.24) **
Value = 6	3.08 (.29) **	1.69 (.21) **	2.33 (.21) **	1.93 (.21) **	2.96 (.25) **
Value = 7	3.70 (.29) **	2.23 (.22) **	5.26 (.42) **	2.21 (.21) **	3.52 (.26) **
Value = 8	4.17 (.30) **	2.56 (.23) **	7.21(1.02) **	2.36 (.22) **	3.85 (.27) **
Value = 9	4.41 (.31) **	2.74 (.24) **		2.44 (.22) **	4.03 (.27) **
Value = 10	4.63 (.31) **	2.93 (.24) **		2.53 (.22) **	4.22 (.28) **
McFadden Pseudo R^2	.023	.006	.019	.021	.020

** . Correlation is significant at the 0.01 level (2-tailed).
* . Correlation is significant at the 0.05 level (2-tailed).

In Model III, Overall Score was used as dependent variable

Table 14-3 Ordinal Logit Regression outcomes relating claim-characteristic-components with PatVal value

In Model I, Forward Citations, Family Size and Opposition were significantly predicting PatVal value. In Model II, only the second component was able to do that. This component relates to the size of the claim set. Model III uses Overall Score instead of PatVal value as a dependent variable. This was done to see the effect of using another value indicator, although Overall Score is of course far less well defined as value predictor. In this model, component 2, 4 and 5 were significant. (The latter two relate to the number of independent claims and the average claim length.) This gave the impression that claims data is better in predicting the value of a composite index of traditional indicators rather than predicting real value estimations.

In the fourth model, both the traditional and claims indicators were used to predict PatVal value. All five claim components turned out to be insignificant. Also the McFadden Pseudo R^2, which is an indication of the explanatory power of the model, diminished slightly from the traditional model. Finally, Model V used only the four variables that proved to be significant in (most of the) the other models. In this model, all four variables were significant but unfortunately the McFadden Pseudo R^2 was still lower than in the model using none of the claims indicators.

So, it seems that the size of the claim set is a significant predictor of patent value. Used together with established indicators as forward citations, family size and opposition, it still managed to contribute to the model. However, the McFadden Pseudo R^2 was lower in this model than in Model I, where only traditional indicators were used.

It is worth the effort to examine if any of the variables underlying component two would perform better as the component itself. In essence, the question here is: can we identify the 'hidden meaning' of component two? Therefore, a new series of regressions were conducted. The results are presented in the table below.

Ordinal Logit Regression

Location	*Model VI*	*Model VII*	*Model VIII*	*Model IX*	*Model X*
Forward Citations Indicator	0.08 (.01) **	0.08 (.01) **	0.08 (.01) **	0.08 (.01) **	0.09 (.01) **
Family Size	0.08 (.01) **	0.08 (.01) **	0.09 (.01) **	0.08 (.01) **	0.09 (.01) **
Opposition	0.23 (0.08) **	0.23 (0.08) **	0.23 (,08) **	0.23 (.08) **	0.24 (.08) **
# of Claims	0.02 (.00) **				
# of Dependent Claims		0.02 (.01) **			
# of Words Total			0.00 (.00)		
# of Words in Dependent Claims Total				0.00 (.00) *	
Ratio of # of Words Total of Independent Claims / All Claims					-0.29 (.19)
Priority Year	-0.01 (.08)	-0.01 (.08)	-0.02 (.08)	-0.02 (.08)	-0.02 (.08)
Technology Field (Instr)	-0.06 (.08)	-0.06 (.08)	-0.08 (.08)	-0.08 (.08)	-0.07 (.08)
Technology Field (ChPha)	-0.27 (.11) *	-0.27 (.11) *	-0.30 (.12) **	-0.30 (.11) **	-0.25 (.11) *
Value					
Value = 1	-1.67 (.19) **	-1.69 (.19) **	-1.77 (.19) **	-1.77 (.19) **	-1.91 (.20) **
Value = 2	-0.38 (.18) *	-0.39 (.18) *	-0.47 (.18) **	-0.47 (.18) **	-0.62 (.19) **
Value = 3	0.58 (.18) **	0.56 (.18) **	0.48 (.18) **	0.48 (.18) **	0.34 (.19)
Value = 4	1.50 (.18) **	1.48 (.18) **	1.40 (.18) **	1.40 (.18) **	1.25 (.19) **
Value = 5	2.34 (.19) **	2.32 (.19) **	2.24 (.18) **	2.24 (.18) **	2.09 (.19) **
Value = 6	3.11 (.20) **	3.09 (.18) **	3.00 (.19) **	3.00 (.19) **	2.86 (.20) **
Value = 7	3.73 (.20) **	3.71 (.20) **	3.62 (.20) **	3.62 (.20) **	3.47 (.21) **
Value = 8	4.20 (.22) **	4.18 (.21) **	4.09 (.21) **	4.10 (.21) **	3.94 (.22) **
Value = 9	4.44 (.22) **	4.43 (.22) **	4.34 (.22) **	4.34 (.22) **	4.19 (.23) **
Value = 10	4.66 (.23) **	4.65 (.23) **	4.56 (.23) **	4.56 (.23) **	4.41 (.23) **
McFadden Pseudo R^2	.024	.024	.023	.023	.023

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Table 14-4 Ordinal Logit Regression outcomes relating claim characteristics with PatVal value

In this series, it has been shown that the number of claims and dependent claims (which are highly correlated) were more significant predictors as the component is. Also the number of words in the dependent claims was significant, but had a zero estimate. The number of words in total became insignificant, although its significance level is .08. To me, this fact, compared with the other tests, is close to being significant to conclude that it cannot be used as a predictor of patent value.

14.4 Assessing marginal effects by using Logistics Regression

One of the major problems using ordinal logit regression is that the value of the estimates can only be interpreted by their sign and not their size. Thus, it teaches us nothing about the size of the effect that the variables have. Therefore, for the models that were able to significantly predict patent value, logistic regression was used in order to extract marginal effects. These can give an impression of the size of the effects.

Logistic regression examines the contribution that an independent variable has on the likelihood of the dependent binary variable to belong to a 'high' or 'low' category. Since logistic regression uses a binary dependent variable, thresholds had to be placed somewhere along the interval of the PatVal value, so that a certain amount of PatVal value forms the 'low value' part and the rest forms the 'high value' part. After logistic regression are calculated, one can use the marginal effect function (which is not present in SPSS, but is available in STATA) to get some sense of the actual effect that the variables have on the dependent variable.

Placing thresholds seems rather arbitrary but was based upon the following rationale: the thresholds should be placed so that the explanatory power of the model is maximized while a reasonably equal distribution of cases among the two groups should be kept. The four claim-set-size variables that were used in this exercise were the 'size component' (extracted from the principal component analysis), the number of claims and dependent claims and the number of words total. Other variables included in the models were forward citations, family size, opposition and the dummy variables for technology field and time. The dependent variable was PatVal value in all models.

For the four different claim variables, each time 10 logit regressions were run (one for each threshold) and their McFadden Pseudo R^2's were compared. Averaging the outcomes over the four variables, the largest average Pseudo R^2's was 0.074 and was found when a threshold was placed at a PatVal value of six. This means that PatVal values 1 to 6 were grouped as 'low patent value' while values 7 to 11 were grouped as 'high patent value'. However, since almost 90% of all patents were now considered to be of 'low' value, it was decided to also use a threshold level of four. For this level, about 65% fell in the 'low value' group which was a more equal distribution. The average Pseudo R^2 at this threshold level for the four models was 0.052.

Table 14-5 on the next page shows the results for the calculation of the marginal effects in three models: that for the size component (Component Two), the number of claims and the number of words total. The model for the number of dependent claims was omitted since it was almost identical with that for the number of claims. The values given in table 14-5 are the marginal effects (dy/dx). For instance, for the number of claims, the marginal effect for the model in which the threshold was set at four, is 0.004. This can be interpreted in the following way: the likelihood of an average patent belonging to the set of 'high valued' patents increases by 0.4% if one claim was added to that patent. For the size component, the marginal effect is extra hard to interpret, since it is not based upon a certain unit.

Logistic Regression, Marginal Effects

(values represent dy/dx)	Model I		Model II		Model III	
	Threshold = 4	Threshold = 6	Threshold = 4	Threshold = 6	Threshold = 4	Threshold = 6
Forward Citations Indicator	0.016 **	0.007 **	0.018 **	0.009 **	0.019 **	0.009 **
Family Size	0.018 **	0.005 **	0.017 **	0.007 **	0.018 **	0.007 **
Opposition	0.064 *	0.023	0.067 **	0.030 **	0.068 **	0.030 **
Component Two	0.026 *	0.011 *				
# of Claims			0.004 **	0.002 **		
# of Words Total					0.0001 **	0.0000
Priority Year	0.010	0.015	0.001	-0.010	0.004 *	-0.009
Technology Field (Instr)	0.048	0.013	0.026	-0.003	0.031	-0.001
Technology Field (ChPha)	0.043	0.059 *	0.075 **	0.032 *	0.088	0.036 *
Mc Fadden Pseudo R^2	.045	.056	.055	.081	.053	.079

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

14-5 Logistic Regression, marginal effects of claim characteristics

The table shows that the marginal effect of the number of claims and especially the number of words total is rather low. In comparing them with the traditional indicators they are much lower, although one has to take into account that the absolute values of the variables also play a role in cross-variable comparisons. The size component performs much better, but it is difficult to interpret its exact meaning. Taking all these considerations into account, these findings strengthen the assumption that using claims is a significant but small contribution next to the traditional indicators.

14.5 The relation of a 'claim set size' indicator to the traditional patent indicators

Finally, it is interesting to see how the claim set size relates to the traditional patent indicators, especially to estimate its independency. Both the component relating to claim-set-size from the principal component analysis as well as the underlying variables were used for this exercise.

It turns out that the size of a claim set was significantly correlated with forward citations, family size and opposition. The number of (dependent) claims showed a strong correlation with forward citations and family size and a remarkably lower correlation with opposition. However, if the number of words was used, these differences became much smaller. Also compare the findings with the correlations between the traditional indicators. Forward citations and family size were about as equally strong correlated with each other as with the number of claims. One can conclude from this that the number of claims is about as independent as an indicator from forward citations or family size as the latter two are independent of each other.

For backward citations and the status indicator, there was no significant correlation with the 'claim size component' or when the number of words was used. When the number of claims was used, there was a small but significant correlation which bears a negative sign. This showed that claims fit better with the more established patent indicators.

Correlation Matrix

	FCI	FS	OP	SI	BCI	CT	#C	#DC	#WT	#WTDC
Forward Citations Indicator (FCI)	1.000									
Family Size (FS)	.220**	1.000								
Opposition (OP)	.087**	.134**	1.000							
Status Indicator (SI)	-.025	-.378**	-.020	1.000						
Backward Citations Indicator (BCI)	.021	-.024	.004	.040*	1.000					
Component Two (CT)	.177**	.163**	.051*	-.019	-.039	1.000				
# of Claims (#C)	.216**	.202**	.066**	-.077**	-.043*	.937**	1.000			
# of Dependent Claims (#DC)	.215**	.192**	.067**	-.070**	-.043**	.933**	.933**	1.000		
# of Words Total (#WT)	.151**	.076**	.104**	.005	-.018	.933**	.779**	.762**	1.000	
# of Words in Dependent Claims Total (#WTDC)	.165**	.107**	.099**	-.002	-.026	.969**	.843**	.852**	.935**	1.000

Table 14-6 Correlation matrix of claim characteristics and traditional indicators

14.6 Conclusion

It seems that the best predictor of patent value based upon full-text is the size of the claim set itself. This is no new knowledge; the literature already mentioned the counting of claims as a way of using claim data to predict patent value. These findings confirm this and add that instead of counting claims, one could count the number of words in the total claim set. However, counting the number of claims outperforms counting the number of words. Also, instead of counting all claims, the number of dependent claims gives generally the same results. When using claim data to predict the Overall Score from PLATO, also the number of independent claims and the average length of claims become significant predictors.

The explanatory power of the models increases hardly by adding claim data and remains around 0.024. The marginal effects of the value estimates were rather low and strengthen the assumption that using claim-set-size as patent indicator is a significant but small contribution to predicting models also using traditional indicators. Further, the size of the claims set showed to be reasonably independent from the traditional indicators.

More sophisticated ways of analyzing claims seem to fail at the moment. It is extremely surprising that the length of independent claims does not give significant value predictions for patent value, since it was one of the most-mentioned aspects by patent attorneys. The findings do not provide a clue as to why this relation was not found.

15 Analysis - Claims as an early value indicator?

For most of the traditional indicators, the reliability of their value predictions increases as the patent ages. For most of them, a patent age of at least five years is required. For business applications, this is often far too late to be of any interest. So, there is a great need to develop indicators that are available early in a patents' lifetime. Fortunately, claims data is available at the publication date and thus could be an interesting indicator in the early years of a patents life. However, patent attorneys were skeptical about the use of applications. They argue that claim sets may still alter during the examination process and in many cases the application is never granted. The next chapter will also examine if the differences between applications and grants themselves can predict patent value.

The PatVal set only includes granted patents. Therefore, it could only be tested if the claim sets of applications can predict patent value, knowing that they will be granted. The characteristics that gave significant estimates for grants were used to test the predicting power of the applications.

The structure of this chapter is somewhat similar to the previous chapter. First, it will be explained how application data was added to the PatVal set. Then, this data is compared to data on the grants. Next, those characteristics that proved to predict patent value in the previous chapter were correlated with the PatVal value and the Overall Score. Since only a small amount of variables was used, a principal component analysis was not needed. Finally, several ordinal logit regressions were used to determine is application data could indeed significantly predict patent value. Marginal effects were not calculated for applications, since they are expected to be similar or even lower to the ones found in the grants.

15.1 Adding applications to the PatVal set

The PatVal dataset had to be expanded with the claim sets of applications. These were extracted from the MicroPatent database, but were incomplete. Claim sets of EP applications entering the EPO via the PCT procedure are only saved as EP applications in MicroPatent if the patent was not filed in English, French or German. Otherwise, the claim sets of applications are saved as PCT applications. Unfortunately, the search query did not include these cases. A second issue was related to language. MicroPatent gives the claim sets of grants in English, French and German. However, for applications only the initial filing language (thus either English, French or German) is included.

For the PatVal set, only 589 applications contain a full English claim set in MicroPatent. To increase the sample size, French and German applications were added. However, the more sophisticated claim characteristics are identified by scripts based upon the English language. Therefore, only basic variables (e.g. number of claims or words) could be used for the French and German claim sets. Fortunately, these were also the best value predictors in case of the grants. A quick analysis looking for possible differences between English, French and German claims revealed that the number of claims was equal for all three, but the number of words differs for the German claims. On average, a total of 450 words were used in a German claim set, while English and French claims contained about 650 words. Therefore, a 'language' dummy was added to the PatVal set.[16] Adding French and German claim

[16] For this dummy, English is used as the base

sets enlarged the PatVal set by respectively 313 and 443 claim sets to a total sample size of 1345 cases.[17] This means that for 50% of all EP grants, the claims of the applications were retrieved.

Before exploring the claim sets of the applications, one final note has to be made. One should realize that all applications in the dataset were eventually granted. However, many applications never make it into a grant (Harhoff et al, 2004, 448). This biases the results. If we assume for the moment that applications that were never granted have lower value than those that do, this means that the claim sets of applications in the PatVal set are already of higher average value than a set of randomly chosen applications.

15.2 Exploring the claim characteristics of applications

In this section, the applications' claim sets are compared with that of the grants. This was done by looking at the number of claims and words (total and average). For the English-written claims, the number of independent and dependent claims was also explored.

The distribution of the number of claims in grants and applications seemed comparable judging from table 15-1. Grants had fewer claims than applications, most likely because some claims are rejected during the examination process. A Wilcoxon Signed Ranks test indicated at a 0.01 significance level (2-tailed) that applications had more claims than their granted counterparts did.

Number of Claims	**Frequency Grants** **N = 2703**	**% Frequency Grants**	**Frequency Applications** **N = 1345**	**% Frequency Applications**
1 - 5	524.00	19.39%	231.00	17.17%
6 - 10	1110.00	41.07%	569.00	42.30%
11 - 15	581.00	21.49%	325.00	24.16%
16 - 20	267.00	9.88%	130.00	9.67%
21 - 25	119.00	4.40%	49.00	3.64%
26 - 50	93.00	3.44%	40.00	2.97%
>50	9.00	0.33%	1.00	0.07%
Mean	10.93		10.69	
Median	9.00		10.00	

Table 15-1 Overview of the PatVal-set – applications and grants compared - number of claims

[17] A last correction that was executed was deleting all cases where the application had only one claim, while the grant had multiple claims. This was done because in a significant amount of claim sets of the applications, erroneously only one claim was given, while the application in fact had more claims.

The number of words total shows that grants contained more words than applications. A Wilcoxon-Signed-Ranks test indicated that applications were indeed smaller in total size than grants at a .01 significance level (2-tailed).

Number of Words Total	Frequency Grants N = 2703	% Frequency Grants	Frequency Applications N = 1345	% Frequency Applications
1-200	119.00	4.40%	96.00	7.14%
201-400	706.00	26.12%	395.00	29.37%
401-600	708.00	26.19%	351.00	26.10%
601-800	511.00	18.90%	237.00	17.62%
801-1000	294.00	10.88%	121.00	9.00%
1001-1500	244.00	9.03%	101.00	7.51%
>1500	121.00	4.48%	44.00	3.27%
Mean	646.46		581.05	
Median	547.00		484.00	

Table 15-2 Overview of the PatVal-set - applications and grants compared - number of claims

Next, the number of words average was examined. Here, it is clear that applications are more represented in the lower regions. The Wilcoxon-Signed-Ranks test confirmed this and showed that in only 5 cases the average length of the grant was longer as its application. Therefore, again at the .01 significance level (2-tailed) it was concluded that the average length of applications is shorter.

Number of Words Average	Frequency Grants N = 2703	% Frequency Grants	Frequency Applications N = 1345	% Frequency Applications
1 - 25	33.00	1.22%	40.00	2.97%
26-50	960.00	35.52%	606.00	45.06%
51-75	1004.00	37.14%	453.00	33.68%
76-100	389.00	14.39%	132.00	9.81%
101-125	178.00	6.59%	66.00	4.91%
126-150	57.00	2.11%	21.00	1.56%
151-200	44.00	1.63%	19.00	1.41%
>200	38.00	1.41%	8.00	0.59%
Mean	66.68		58.99	
Median	57.00		52.00	

Table 15-3 Overview of the PatVal-set - applications and grants compared - number of words average

In addition, also the difference between the independent claims and dependent claims in applications and grants was examined. However, this could only be done for cases with an English claim. Table 15-4 on the next page shows that applications tend to have more independent claims and just slightly more dependent claims. Grants had much longer independent claims. The difference in length for dependent claims is not large. A Wilcoxon-Signed-Ranks test confirmed that there was a difference in all four variables at the .01 significance level (2-tailed).

	Number of Independent Claims		Number of Words Average in Independent Claims		Number of Dependent Claims		Number of Words Average in Dependent Claims	
	Grant N = 589	Application N = 589	Grant N = 589	Application N = 589	Grant N = 589	Application N = 589	Grant N = 589	Application N = 589
Mean	1.43	2.00	170.31	126.56	8.40	9.11	48.99	46.99
Median	1.00	2.00	156.00	108.33	7.00	8.00	42.96	42.63
Std Dev	0.80	1.38	94.68	83.30	5.37	5.63	28.12	20.39
Min	1.00	1.00	23.00	15.00	1.00	1.00	14.50	15.33
Max	8.00	13.00	839.00	841.00	34.00	42.00	435.00	165.33

Table 15-4 Overview of the PatVal-set - applications and grants compared - other characteristics

15.3 The correlation of claim characteristics of applications with value

As a next step, the claim characteristics mentioned above were tested on their ability to indicate value by constructing a correlation matrix, using again PatVal value and Overall Score as value indicators.

Correlation Matrix

	PatVal Value	Overall Score
# of Claims	.109 **	.209 **
# of Words Total	.067 *	.140 **
# of Words Average	-.015	-.047
	(N = 1100)	*(N = 1345)*
# of Independent Claims	-.025	.095 *
# of Dependent Claims	0.10 *	0.16 **
# of Words in Independent Claims Average	-.066	-.053
# of Words in Dependent Claims Average	-0.10 *	-0.14 **
	(N = 481)	*(N = 589)*

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Table 15-5 Correlation matrix of claim characteristics using claim sets of applications with PatVal value and Overall Score

The table shows that the size of the claim set of applications has predicting power for value, just as for grants. Both the number of claims and number of words total were significantly correlated with PatVal value and the Overall Score. Do note that the number of claims in applications was .895 correlated at the .01 significance level to the number of claims in grants. The same holds for the total number of words in applications, which was .886 correlated at the .01 significance level to the total number of words in grants. The number of words average did not give a significant correlation with PatVal value or Overall Score.

As for independent claims in the applications, there is a significant relation found between their number and the Overall Score. The number of dependent claims was also positively correlated with PatVal Value and the Overall Score, which is in line with the findings for grants. Interestingly, the average length of dependent claims was correlated significantly to both value indicators.

15.4 Ordinal Logit Regression using application data

Ordinal logit regressions were run to test which of the claim characteristics that correlate with patent value could indeed predict it. Only Family Size is used as a traditional indicator. This is because Family Size is the only of the three established traditional indicators not suffering from timeliness. Since this exercise is meant to examine the possibility of developing an 'early' indicator, there was no point in testing its contribution compared to traditional indicators which become reliable long after publication.

Ordinal Logit Regression

Location	*Model XI*	*Model XII*	*Model XIII*	*Model XIV*	*Model XV*
Family Size	.10 (.01) **	0.10 (.02) **	0.10 (.02) **	0.08 (.03) **	0.08 (.03) **
# of Claims		0.02 (.01) **			
# of Words Total			0.00 (.00)		
# of Dependent Claims				0.02 (.01) *	
# of Words in Dependent Claims Average					0.00 (.00)
Priority Year	0.02 (.08)	0.10 (.12)	0.09 (.12)	0.03 (.17)	0.06 (.18)
Technology Field (Instr)	-0.10 (.08)	0.06 (.12)	0.03 (.12)	-0.24 (.17)	-0.28 (.19)
Technology Field (ChPha)	-0.37 (.11) **	-0.50 (.17) **	-0.54 (.18) **	-0.65 (.28) *	-0.71 (.31) *
Language (French) (1)	0.13 (.13)	0.19 (.15)	0.20 (.15)	xxx	xxx
Language (German) (1)	.25 (.10) **	0.31 (.12) **	0.29 (.13) *	xxx	xxx
Value					
Value = 1	-1.88 (.22) **	-1.44 (.32) **	-1.61 (.31) **	-2.56 (.44) **	-3.03 (.49) **
Value = 2	-0.59 (.22) **	-0.17 (.31)	-0.34 (.30)	-1.28 (.42) **	-1.76 (.47) **
Value = 3	0.35 (.22)	0.71 (.31) *	0.54 (.30)	-0.34 (.42)	-0.81 (.47)
Value = 4	1.25 (.22) **	1.59 (.31) **	1.42 (.30) **	0.63 (.42)	0.11 (.47)
Value = 5	2.08 (.22) **	2.52 (.32) **	2.34 (.31) **	1.59 (.43) **	1.11 (.47) *
Value = 6	2.82 (.23) **	3.24 (.33) **	3.06 (.32) **	2.54 (.44) **	2.02 (.49) **
Value = 7	3.43 (.23) **	3.79 (.34) **	3.60 (.33) **	3.46 (.48) **	2.86 (.52) **
Value = 8	3.89 (.24) **	3.99 (.35) **	3.81 (.34) **	3.91 (.51) **	3.31 (.55) **
Value = 9	4.13 (.25) **	4.14 (.35) **	3.96 (.34) **	4.37 (.56) **	3.77 (.59) **
Value = 10	4.34 (.26) **	4.25 (.36) **	4.06 (.34) **	5.23 (.71) **	4.63 (.74) **
McFadden Pseudo R^2	.017	.017	.016	.015	.015

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

For Model XI, N = 2275, for Models XII and XIII, N = 1100, for Model XIV, N = 541, for Model XV, N = 477
(1) Estimates could sometimes not be calculated since the dummies were considered to be redundant

Table 15-6 Ordinal Logit Regression outcomes relating claim characteristics using claim sets of applications with PatVal value

In all models, family size has a significant estimate. The number of claims was again a significant predictor of patent value, although its estimate is very low. Also the number of dependent claims was a significant estimate for patent value. The estimates were positive in both cases, suggesting that a larger claim set predicts more value. The number of words total has an insignificant estimate, although removing the language dummies made the estimate significant. This indicated that there are indeed differences in the number of words caused by the language that is used.

The explanatory power of the models hardly increased by using claim data if family size was already included in the model. If the number of claims is used solely in the model, the explanatory power of the model is 0.007 with a significant estimate at the .01 level (2-tailed). This means that by itself, the claim set size in applications has still some explanatory power, but if family size is also included in the model, its role diminishes.

15.5 Conclusion

To conclude, it can be said that the regressions show that size of the claim set, represented by the number of (dependent) claims from applications, can predict patent value to some extent. Therefore, the size of the claim set in applications can be used as an early value predictor. The explanatory power of the model is about a third of the model using grants. However, since forward citations and opposition are excluded, this is still rather well.

16 Analysis - Differences between applications and grants as value indicator

So far, grants and applications were used separately as unit of analysis. However, adding applications to the dataset made it feasible to examine if the differences between applications and grants can predict patent value. The interviews with patent attorneys led to two contradicting hypothesis. On the one hand, it was expected that differences indicate that the desired patent was not approved and on top of this, its scope was narrowed down. Thus, differences would lead to lower patent value. On the other hand, differences could indicate a maximized scope which was pushed especially for those patents that are expected to be of higher value. Thus, differences would signal on average patents of more value.

This chapter tries to empirically determine which of the two hypotheses is correct. In addition, it will be discussed if Backward-X-citations can predict differences, since theoretically their presence in a search report indicates that the applicant has to change (a part of) the claim set. If this can be empirically validated, this knowledge could improve the accuracy of a claims indicator based upon applications, if used in a combined fashion.

This chapter takes of exploring the characteristics of the differences between applications and grants. Then, a correlation matrix is presented showing what kind of differences correlate with PatVal value, the Overall Score and the number of backward-X-citations. From the variables that showed significant correlation with PatVal value and/or the Overall Score, a principal component analysis extracted two components. These were used in ordinal logit regressions, together with some 'raw' variables, to examine if they could significantly predict patent value. For those variables that succeeded, it was tested if they correlated with the size of the claim set, which would bias these findings. Next, the marginal effects of the successful value predictors were calculated to examine their effect. Further, the difference indicators were compared with the traditional patent indicators to determine their independence. Finally, the chapter ends with a short attempt to determine if differences can be predicted by using backward-X-citations.

16.1 Exploring the 'difference' characteristics

For six claim characteristics, the differences were calculated. This was done by subtracting the application value from the grant value. An overview can be found in the table below. At first, the number of words in independent claims average was also included, but it provided no significant correlations whatsoever. Therefore, it was left out of the analyses.

(for all calculations: Grants minus Applications)	**Difference in number of claims** N =1345	**Difference in number of independent claims** N = 589	**Difference in number of dependent claims** N = 589	**Difference in number of words total** N = 1345	**Difference in number of words average** N = 1345	**Difference in the number of words in independent claims total** N = 589
Mean	-0.97	-0.57	-0.71	28.61	11.74	7.59
Median	0.00	0.00	0.00	25.00	6.60	11.00
Minimum	-23.00	-12.00	-21.00	-1445.00	-34.86	-731.00
Maximum	22.00	4.00	20.00	1712.00	376.00	761.00
Standard Deviation	2.78	1.22	3.18	178.50	23.12	138.81

Table 16-1 Overview of the PatVal-set - differences between applications and grants

For more insight, the size of the differences for the number of claims was explored. Almost half of the patents in the set have no change in their number of claims. Of thirty percent of the claim sets, one to five claims is rejected during the examination procedure. Also, five percent of the applications gains claims. The most likely explanation for such cases is that a claim with a broad scope was not accepted and was reformulated in several claims with more specific scopes. However, the combined scope of these claims remains smaller as that of the initial proposed claim. When this exercise for the number of words was completed, it turned out that only 40 out of 1345 patents had no changes in the number of words while 70% had a change of more than 25 words.

(for all calculations: Grants minus Applications)	**Size of difference in number of claims between grants and applications, frequency** N =1345	**Frequency %**
Greater than -10	22.00	1.64%
-10 to -6	51.00	3.79%
-5 to -1	418.00	31.08%
0	783.00	58.22%
+1 to +5	61.00	4.54%
Greater than +5	10.00	0.74%

Table 16-2 Overview of the PatVal-set – the size of the differences in number of claims

16.2 The correlation of several difference calculations with value

The correlation matrix in table 16-3 presents the relations to value for the difference variables. In addition, the correlation between the number of X-citations that an application received and the differences was tested. [18] An X-citation is given by an examiner if the application holds invalid claims, either because they are not new or their inventive step is considered to be too small. A relation with the differences was expected, since an X-citation indicates that the applicant has to drop or alter a claim or even several claims, before the patent can be granted. The number of X-citations was extracted from PLATO.

[18] For more information on X- and Y-citations, see section 20.1.2 on Backward Citations. The correlation was also examined for Y-citations, but this delivered no significant correlations for all variables. Therefore, these results are not presented here.

Correlation Matrix

In each of the following rows, the variables are the differences between grants and applications.	**PatVal Value** N = 1100	**Overall Score** N = 1345	**# X-Citations Received** N =1345
# of Claims	-.060 *	-.051	-.081 **
Absolute (# of Claims)	.060 *	.071 **	.089 **
# of Claims (only negative differences)	-.065 *	-.068 *	-.094 **
# of Independent Claims (1)	-.024	-.077 *	.044
Absolute (# of Independent Claims) (1)	-.034	.083 *	.012
# of Independent Claims (only negative differences) (1)	.004	-.089 *	.019
# of Dependent Claims (2)	-0.09 *	0.01	-0.06
Absolute (# of Dependent Claims) (2)	0.10 *	0.07	0.09 *
# of Dependent Claims (only negative differences) (2)	-.110 *	-.033	-.092 *
# of Words Total, relative difference	-.117 **	-.132 **	-.044
Absolute (# of Words Total), relative difference	.001	-.062 *	.047
# of Words Total, relative difference (only negative differences)	-.096 **	-.070 **	-.065 *
# of Words Average, relative difference	-.073 *	-.107 **	.067 *
Absolute (# of Words Average), relative difference	-.071 *	-.117 **	068 *
# of Words Average, relative difference (only negative differences)	-.024	-.003	.013
# of Words Total in Independent Claims, relative difference (1)	-.034	-.055	.112 **
Absolute (# of Words Total in Independent Claims), relative difference (1)	-.077	.008	.106 **
# of Words Total in Independent Claims, relative difference (only negative differences) (1)	.027	-.052	.028

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

(1) Respectively, N = 541; N = 661 and N = 661

(2) Respectively, N = 481, N = 589 and N = 589

Table 16-3 Correlation matrix of differences between grants and applications with PatVal value and Overall Score

As the table above shows, several types of difference-calculations were used to see which one gave the best fit. When the number of claims was the unit of analysis, most significant results were booked if all cases with positive differences were transformed into zero's. So, only negative and zero differences were included. If the number of words was used, the relative difference gave the best results. These were calculated by dividing the differences by the sum of the application and the grant. Do note that for the number of claims, the variable holds a negative sign, while for the number of words, both positive and negative values are present.

The difference in the number of claims correlated with patent value in most of the cases. Since this variable is negative, the correlation actually states that more differences indicate higher patent value. For independent claims, correlation was only found with the Overall Score. The relative difference in the total number of words has a negative significant correlation with both PatVal value and Overall Score. Here, the findings are more difficult to interpret since about 35% of the cases of the variable is negative while 65% of the cases is positive. On average, the findings show that if more words are added to the grant, this is of negative influence on its value.

Finally, among the more interesting correlations with value, there was a significant relation between the relative differences in the average number of words, both in the normal and absolute case. There seemed to be no significant correlation for the number of total words in the independent claims with the value indicators.

When looking for characteristics correlating with the number of X-citations, two characteristics jumped out. First, the difference in the number of claims was significant and negative. Again, this indicates actually that more differences signal more X-citations. Second, the number of words in the independent claims total was also correlated to the number of X-citations in two of the three cases. This indicated that changes are made primarily in the independent claims.

16.3 Principal Component Analysis using 'differences' data

A principal component analysis was conducted to see if any of the variables from the correlation matrix were interrelated. Two components were extracted from this exercise, as can be seen below in table 16-4. The first components relates to size of the differences in the total claim set. The second components points towards changes in the number of independent claims. Together, they explained more than three-quarter of the variance. It is unclear what the role of the differences in total number of words was, since it loads considerably on both components.

Rotated Component Matrix[a]

	Component	
	1	**2**
# of Claims (only negative differences)	.980	.099
# of Independent Claims (only negative differences)	.211	.706
# of Dependent Claims (only negative differences)	.970	-.073
# of Words Total, relative difference	***.651***	***.555***
# of Words Average, relative difference	-.429	.663
# of Words Total in Independent Claims, relative difference	.061	.900

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 3 iterations.

	Initial Eigenvalues		
Component	Total	% of Variance	Cumulative %
1	2.71	45.11	45.11
2	1.922	32.03	77.14

Extraction Method: Principal Component Analysis.

Table 16-4 Principal Component Analysis outcomes of differences characteristics

16.4 Ordinal Logit Regression using 'differences' data

The next step in this exercise is to run a series of ordinal logit regressions with these components. The results are presented on the next page. Component one gave a significant estimate with a negative sign. Next, it was tested if the underlying variables could be used as value predictors. The models XVII and XVIII showed this was the case. In all three models, the estimate has a negative sign. For models XVII and XVIII, this indicates that the more differences there are, the higher the value of a patent is expected to be. As for model XIX, the model predicts on average that more changes in words signals lower patent value. However, if the number of words declines, this effect reverses.

Interestingly, the explanatory power of the model when using the difference in dependent claims rose with about 20% compared to the model using only traditional indicators. So, it seems that a claims indicator based upon differences outperforms one based upon the size of the claim set. If the difference in total number of words was used, the estimate was not significant but came close to being so. Therefore, it is believed that the differences in number of words total could still be related to patent value.

Ordinal Logit Regression

Location	Model XVI	Model XVII	Model XVIII	Model XIX
Forward Citations Indicator	0.11 (.03) **	0.08 (.02) **	0.11 (.03) **	0.08 (.02) **
Family Size	0.07 (.03) *	0.08 (.02) **	0.07 (.03) *	0.08 (.02) **
Opposition	0.43 (.28)	0.46 (.14) **	0.42 (.28)	0.47 (.14) **
# Claims (only negative differences)		-0.05 (.02) *		
# Dep. Claims (only negative differences)			-0.07 (.03) *	
# Words Total (relative difference) (1)				-0.41 (.23)
Difference Component One	-0.16 (.08) *			
Difference Component Two	0.08 (.08)			
Priority Year	-0.03 (.18)	0.08 (.12)	-0.03 (.18)	0.08 (.12)
Technology Field (Instr)	-0.34 (.19)	0.05 (.12)	-0.33 (.19)	0.05 (.12)
Technology Field (ChPha)	-0.74 (.30) *	-0.41 (.18) *	-0.71 (.30) *	-0.37 (.18) *
Language (French) (2)	…	0.16 (.15)	…	0.19 (.15)
Language (German) (2)	…	0.26 (.12) *	…	0.18 (.14)
Value				
Value = 1	-2.60 (.49) **	-1.37 (.31) **	-2.49 (.48) **	-1.43 (.32) **
Value = 2	-1.30 (.47) **	-0.09 (.31)	-1.19 (.47) **	-0.15 (.31)
Value = 3	-0.31 (.47)	0.81 (.31) **	-0.21 (.46)	0.74 (.31) *
Value = 4	0.65 (.47)	1.70 (.31) **	0.75 (.46)	1.63 (.31) **
Value = 5	1.67 (.47) **	2.65 (.32) **	1.78 (.47) **	2.58 (.32) **
Value = 6	2.61 (.49) **	3.39 (.33) **	2.72 (.49) **	3.32 (.33) **
Value = 7	3.47 (.52) **	3.94 (.34) **	3.58 (.52) **	3.87 (.34) **
Value = 8	3.92 (.55) **	4.15 (.34) **	4.03 (.55) **	4.08 (.35) **
Value = 9	4.39 (.60) **	4.30 (.35) **	4.50 (.60) **	4.23 (.35) **
Value = 10	5.26 (.74) **	4.41 (.35) **	5.37 (.74) **	4.34 (.35) **
McFadden Pseudo R^2	.028	.024	.028	.024

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

(1) The significance level was .075
(2) In models XVI and XVIII, the language dummies were considered redundant. Instead, dots (…) were printed.

Table 16-5 Ordinal Logit Regression outcomes relating differences characteristics with PatVal value

Also, consider the following graph for additional understanding. For each level of PatVal value, the relative number of patents that change or not change in their number of claims is given. Thus, the two lines are directly related: if one increases, the other decreases. It is striking to see that for the lower PatVal values, the relative number of changes is lower as for the higher PatVal values. Thus, changes are more likey to occur for higher PatVal values.

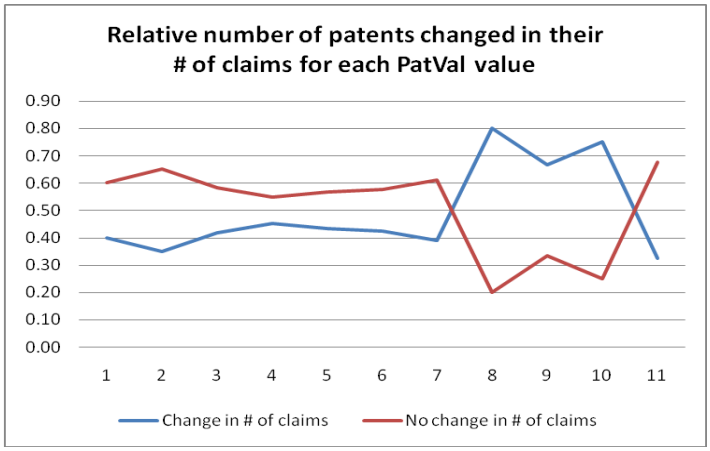

Figure 16-1 Relative number of patents changed in their # of claims for each PatVal value

To be perfectly sure that the size of the claim set is not influencing the findings for the differences, the correlation between the two was tested. Table 16-6 presents a correlation matrix showing that the correlation between the size of the claim set and the differences between applications and grants is only little, although significant. This fact, combined with the fact that the explanatory power of the model increases when using differences, leads to the conclusion that the size of the claim set and the differences between applications and grants are to independent value indicators.

Correlation Matrix

	#C	#WT	DCO	D#C	D#DC	D#WT
# Claims (#C)	1.000					
# Words Total (#WT)	.779 **	1.000				
Difference ComponentOne (DCO)	.121 **	.051	1.000			
Dif # Claims (only neg. dif.) (D#C)	.055 *	.013	.980 **	1.000		
Dif # Dep. Claims (only neg. dif.) (D#DC)	.053	-.003	.970 **	.960 **	1.000	
Dif. # Words Total (rel. dif.) (D#WT)	.114 **	.110 **	.651 **	.620 **	.575 **	1.000

Table 16-6 Correlation matrix of difference characteristics and the size of the claim set

16.5 Assessing marginal effects by using Logistics Regression on 'differences' data

The newly extracted variables that are based upon the differences between applications and grants will also be tested on the marginal effect they have. In the logistic regressions that were conducted to calculate these marginal effects, a threshold of two was used. For higher thresholds, it turned out that the variables relating to the differences became insignificant. For a threshold at two, about 25% of the patents belongs to the 'low valued' group of patents, while 75% belongs to the 'high valued' group. This suggests that differences are more equipped to determine whether or not a patent has at least some value, as opposed to discriminate between highly valued and extremely highly valued patents. The McFadden Pseudo R^2 for the models was, on average, 0.05.

Three differences variables were used: the difference component one, the difference in number of claims and the difference in number of words total. A model for dependent claims was again excluded because it was almost identical with that of the number of claims. Other variables included in the models were forward citations, family size, opposition and the dummy variables for technology field, time and language. The dependent variable was PatVal value in all models.

The table below presents the marginal effects for these three models. The findings can be interpreted as followed: for a given variable, say the difference in number of claims, the likelihood of an average patent belonging to the set of 'high valued' patents increases by 1.9% if one more claim was dropped during the examination procedure. Again, these findings are harder to interpret for the difference component or for the model based upon the differences in number of words.

Logistic Regression, Marginal Effects

(values represent dy/dx)	Model I Threshold = 2	Model II Threshold = 2	Model III Threshold = 2
Forward Citations Indicator	0.018 **	0.011 *	0.012 *
Family Size	0.018 *	0.021 **	0.021 **
Opposition	0.091	0.044	0.047
Difference ComponentOne	-0.055 *		
Difference # of Claims		-0.019 **	
Difference # of Words Total			-0.147 **
Priority Year	0.024	-0.006	-0.006
Technology Field (Instr)	0.056	-0.028	-0.027
Technology Field (ChPha)	0.117 *	0.022	0.009
Language (French) (1)	(…)	-0.110 **	-0.123 **
Language (German) (1)	(…)	-0.066 *	-0.038
Mc Fadden Pseudo R^2	.068	.042	.040

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

(1) The language dummies were dropped in model I because of collinearity. Instead, dots (…) were printed.

16-7 Logistic Regression, marginal effects of differences

Table 16-7 above shows considerably larger marginal effects compared to the ones found by using the claim set size. Especially the differences in the number of words total has a high marginal effect, but one has to take into consideration that this variable is constructed by the difference divided by the total amount of words in grants and applications. This makes these findings more logical. When compared to the traditional indicators, the differences for the number of claims (which is the only variable with a similar unit of analysis as the traditional indicators) have on average a higher marginal effect. All in all, this shows that the differences are indeed a rather powerful indicator with a positive effect on patent value.

16.6 The relation of a 'difference' indicator to the traditional patent indicators

It is checked with a correlation matrix to which traditional indicators the findings in this chapter correlate. Also the number of X-citations was used here. The results are shown in the table below. The difference indicators seem only to correlate with the forward citations indicator and the number of X-citations. The correlation with forward citations shows that the value estimations made by differences are in line that of the most established traditional indicator. The correlation with the X-citations shows that the more X-citations are received, the more likely a higher amount of differences in number of (dependent) claims is.

Correlation Matrix

	FCI	FS	OP	SI	BCI	#XCR	DCO	D#C	D#DC
Forward Citations Indicator (FCI)	1.000								
Family Size (FS)	.220 **	1.000							
Opposition (OP)	.087 **	.134 **	1.000						
Status Indicator (SI)	-.025	-.378 **	-.020	1.000					
Backward Citations Indicator (BCI)	.021	-.024	.004	.040 *	1.000				
# X citations received (#XCR)	.082 **	.024	.001	.002	-.022	1.000			
Difference Component One (DCO)	-.090 **	.024	-.054	-.027	.063	-.086 *	1.000		
Difference # Claims (only negative differences) (D#C)	-.075 **	-.037	.007	-.051	.022	-.094 **	.980 **	1.000	
Difference # Dep. Claims (only negative differences) (D#DC)	-.087 **	.032	-.039	-.026	.067	-.080 *	.970 **	.960 **	1.000

Table 16-8 Correlation matrix of the difference characteristics and traditional indicators

16.7 Predicting differences by X-citations?

The final part of this chapter will briefly examine if differences can be predicted by Backward-X-citations. As mentioned before, the theory states that during the examination procedure of a patent application, the examiner presents a search report. The report includes a list of relevant backward citations. For patents in the EPO, the citations are always given a 'letter'. The letter 'X' refers to backward citations that have prior art showing that the novelty or inventive step criteria is not met by the application. This makes it unacceptable for the examiner to grant (a part of) the claim set in that format. Thus, either the applicant should drop its application or he/she should make changes, which in most cases comes down to narrow down the scope of the patent.

Such knowledge is very useful to support value predictions based upon claim data in applications. Since the claim sets of applications can change during the examination procedure, value predictions based upon the claim set change. If the likelihood of changes can be entered into the model, this might enhance the accuracy of predictions for datasets containing patents that are not yet granted.

The correlation matrix in table 16-8 already showed that the number of differences in claims and the differences in the number of words total in independent claims are significantly correlated with X-citations. This suggests that

when the number of X-citations increases, the likelihood of changes is higher. Unfortunately, the dataset includes just 46 cases with X-citations and a claim set for applications. This is too small for any serious analysis. The only interesting thing that could be identified was the percentage of zero differences for these 46 cases. For the difference in the number of claims, only 28% of these cases had zero differences. For the differences in the number of words total in independent claims, none of the cases had zero changes. But since these percentages for the entire dataset (thus with or without having an X-citation) are respectively 29% and 2%, one can clearly not state there is any significant difference.

16.8 Conclusion

This chapter showed that most patents have a change in the number of words when going through the examination procedure and about 50% also has a change in its number of claims. The difference in the number of claims between grants and applications proved to be the best predictor for patent value: since the estimate has a negative sign, the differences signal higher patent value. This suggests the rejection of the hypothesis that differences indicate a smaller scope and thus lower patent value. Instead, the findings are more supportive of the hypothesis that differences indicate an aim for a maximized scope and/or extra effort of the applicant which signals higher expected patent value.

The explanatory power of a model using the differences and the traditional indicators is considerably higher than models using only the traditional indicators. The differences are only slightly correlated with the size of the claim set, which means that the findings are not influenced by the sheer size of the claim set. When looking at the marginal effects, they are on average higher as the ones found for the size of the claim set. This strengthens the findings that differences between grants and applications provide a claims indicator with a real contribution to the already existing indicators. Further, the differences showed hardly any correlation with the traditional indicators which makes differences a rather independent indicator. Only for forward citations significant correlations were found. Finally, an attempt was made to relate backward-X-citations to the number of differences. However, the datasets are not equipped to answer this question.

17 Conclusion

Due to the rise of importance of IP and the unique features patent data has, statistical patent valuation is increasingly used to identify patents that are valued 'higher-than-average'. Scientific literature has shown many applications for statistical patent valuation as for instance analysis of technological change or measurement of innovativeness. Patent attorneys in IP&S, Philips, believe that statistical patent valuation can be used to support the scouting process, to analyze and compare patent portfolios of competitors with their own portfolios and to assist in M&A matters.

In order to perform statistical patent valuation, several patent indicators were developed which use patent metadata (e.g. forward citations, renewal data). Because most of them suffer from timeliness and fail to include content-related information, Reitzig (2004a) proposed to construct a full-text patent indicator. Developing such an indicator was the goal of this thesis project. Therefore, the following research question was put forward:

How can full-text patent data be used to predict the value of a patent?

Three sub-research questions were formulated. These will be answered here, after which the main research question is answered. The answer on the first sub-research question will provide us with an overview of the claim characteristics that were identified and the external factors that are of influence. In general, it covers the chapters 11 and 13. The second sub-research question answers the question if these claim characteristics are indeed capable of predicting patent value. By doing this, it covers the findings of chapters 14 to 16. Finally, the answer on the third sub-research question gives an indication to what extent a claims indicator is a contribution next to the existing patent indicators. It covers findings from the same chapters as sub-research question two.

After answering the three sub-research questions, the main research question will be answered, which provides the most basic overview of the research findings. Together, these answers form the conclusion of this thesis project.

17.1 Answering the sub-research questions

1) *Which characteristics of the full-text of a patent could be related with value? And what external factors may disturb these relations?*

The term 'full-text' was narrowed down to the claim set of a patent. The claim set forms the legal heart of a patent and consists of a well-structured set of claims: single juridical-written sentences describing the invention that is claimed. Ninety percent of all patens have between 1 and 20 claims. In interviews with patent attorneys, in which the guiding question ran "*what claim characteristics could hint you that a patent might be of higher-than-average value?"* several claim characteristics were identified that could indicate patent value.

The sheer number of claims was believed to be positively related to patent value. Each claim represents an inventive contribution by itself, thus more claims signal a wider field of application and thus higher patent value. Also, applicants invested more effort in constructing larger claim sets which signals higher expected value. A third rationale is that extra claims require extra fees been paid. Finally, additional dependent claims provide more fall-back positions strengthening the legal position of the patent.

Claims can be divided into independent and dependent claims. About 70% of all patents have just one independent claim, while just 5% contains three or more independent claims. Independent claims form the 'spine' on which the rest of the claim set is build. Dependent claims always refer to independent claims and indicate more

specific applications of the invention that was described in the independent claim. On average, independent claims contain 150 words while dependent claims contain generally about 50 words. The length of the independent claims was believed to be negatively related to patent value, as longer claims tend to narrow a patent's scope.

Independent claims could be divided into product and method claims. Product claims have a wider legal protection and could therefore be of more value. Also, the formulation of independent claims could be of influence on patent value. If formulated in a functional setting instead of a 'portrait' setting, the scope of a patent could be wider and thus the patent might be of more value. If a claim includes many so-called 'limiting words', this can indicate a more narrow scope and thus less patent value is expected. Finally, the ratio between the length of the preamble and characterizing portion could point towards value.

The second part of the sub-research question questioned if external factors might be of influence on the relation between the claim characteristics and patent value. There were several external factors identified as being influential on the claim set. For three of them, correction was not feasible or needed. Language differences were not controlled for since only English claim sets were used. The only exception to this was the use of French and German applications but at that moment, a language dummy was introduced. The height of claim-related fees of the EPO and USPTO did not change much until very recently and thus for this project, they were of no influence. The applicant's writing style and expertise is a factor of influence, but could not be corrected for.

Three other factors were of influence. Country differences were identified by comparing EP and US claim sets. US patents have on average 4.5 claims more than EP patents. Also, US patents contain on average one independent claim more. Also, their independent claims are generally 10 words shorter as in their EP counterparts. Throughout time, claim sets also increased in volume. They gained on average four claims and half an independent claim. Also, the length of claims seems to decrease somewhat over time. Finally, claim sets seem to differ in different fields of technology. Although the statistical evidence did provide convincing evidence for this assumption (only the number of independent claims and average claim length significantly differ per technology field), the back-up of qualitative data (patent attorneys expected that the type of technology influences the claim set) led to the conclusion that controlling for technology fields was necessary.

As a final remark, attorneys stressed that it is important to distinguish the claim sets of applications and that of grants. During the examination procedure, claim sets may alter which makes it tricky to use the claims in applications. About 40% of the patents in the PatVal set did not change in the number of claims during the process but 70% of them had a significant amount of words that changed. About 75% of the patents changing in number of claims looses one to five claims, while 15% of them gains claims. The number of independent claims in grants compared to applications drops by 25%, while their length increases by almost the same percentage. Thus, applications have more and shorter independent claims.

As for the differences between applications and grants, two hypotheses were formulated regarding their relation with patent value. One argues that the amount of differences between applications and grants could signal to what extent the desired patent was approved by the examiner. Since examiners narrow down the scope of a patent and scope size is believed to be related to patent value, the difference rate could signal less patent value. The other hypothesis states that claim sets that have been changed are more likely to have maximized their scope, while at the same time they signal more investment of the applicant in the patent which signals higher expected patent value. The differences between applications and grants also showed a significant correlation with the number of backward-X-citations, which were identified as a possible predictor of these differences.

2) Which of the identified characteristics are significantly related to patent value?

Claim sets of grants were used to determine which characteristics can predict patent value. The best predictor was the size of the claim set. This can be represented by either the number of claims or the number of total words in a patent. However counting claims outperforms counting words. Instead of all claims, only the dependent claims can be used without loss of explanatory power of the model. With principal component analysis, also a component based upon number of (dependent) claims and words was constructed which was significantly predicting patent value. This means that also sort of 'composite claims indicator' can be used as predictor of patent value.

Surprisingly, the length of independent claims was not a significant predictor of patent value, although it showed a negative correlation with patent value. Since the data for the length of independent claims is rather reliable, it is not expected that more accurate data construction would change these findings. Therefore, an explanation for this result cannot be given at the moment. Some patent attorneys suggested that the personal writing style of applicants might be of influence here. Perhaps some attorneys are better capable of writing long claims that still have a broad scope than others, which biases the relation between independent claim length and patent value. Also, the patent attorneys mentioned that maturity of technology fields could be of influence on the average length of independent claims.

The number of independent claims and average claim length could not predict patent value but was significant in predicting the Overall Score of PLATO. All other claim characteristics did not show ability to predict patent value. The number of limiting words might be too much dependent on the style of the applicant, since only a limited set of words was tested. The degree of functionality of claims showed no correlation with patent value. This can be caused by the fact that functionality was not measured in a successful way.

When the claim sets of applications were used instead of grants, the size of the claim set remains a significant predictor of patent value. However, the results were less convincing as they were for grants.

The differences between grants and applications are generally positively correlated with patent value. The difference in volume of the claim set, measured by the number of (dependent) claims, is a significant predictor of patent value. Also a component based upon several difference variables (which was extracted by principal component analysis) showed to have significant predicting power for patent value. The component also seemed to relate to the sheer change in volume of the claim set. Simplified, these findings come down to the following: if claims are dropped during the examination procedure, it becomes more likely that the patent will have a higher value. This suggests two things. First of all, the hypothesis that differences indicate maximized claim scope and/or higher expected patent value by the applicant (shown by his/her extra invested effort or more developed markets) is more or less confirmed. Second, the fact that the scope of a patent is narrowed down during the examination procedure seems to have a smaller effect on the expected value of a patent than was expected beforehand.

Finally, a positive significant correlation was found between the amount of differences in claim sets and the number of backward-X-citations a patent receives. This could mean that X-citations might be used as a predictor for changes in the claim set during the examination procedure. Such knowledge can enhance the accuracy of an early value prediction for applications that have not yet been granted.

3) To what extent does a full-text indicator improve the statistical methods to analyze patent value?

The size of the claim set was used in a same model with the established traditional patent indicators forward citations, family size and opposition. All four variables were significant and thus the size of the claim set makes a contribution to these three traditional indicators. However, the explanatory power of the model rises by only a

fraction. A model using only traditional indicators has a McFadden Pseudo R^2 of 0.023. The calculated marginal effects confirm that the contribution of the size of the claim set is significant but small. When data on the number of (dependent) claims is added, it rises to 0.024. More improvement was made if the differences between applications and grants were put in the model. Here, the R^2 rose to 0.028 in models using the number of dependent claims or a component relating to the change in volume of the claim set. Also the marginal effects suggest that the contribution of differences is larger than it was for the size of the claim set. Thus, a claims indicator based upon differences between grants and applications outperforms one based upon the size of the granted claim set.

The models based upon application data have an even lower explanatory power, which circles around 0.016 but is primarily caused by the inclusion of family size. Here, the traditional indicators forward citations and opposition were omitted since the models were testing the ability of early value recognition (forward citations and opposition suffer of timeliness). A model using solely claim data from applications has a McFadden Pseudo R^2 of 0.007.

In general, it can be concluded that the use of claims improves the current prediction models in three ways. First, including the size of the claim set next to the traditional indicators in prediction models enhances these models, although the explanatory power hardly increases. Second, using the amount of differences in number of claims between grants and applications next to the traditional indicators in prediction models again enhances these models, but now the explanatory power increases significantly. Third, the claim set of applications can be used as an early value indication. This makes claim data indeed an interesting and welcome addition to the set of patent indicators that exist.

17.2 Answering the research question

The main research question of this thesis project will now be answered.

How can full-text patent data be used to predict the value of a patent?

The best way to predict patent value by using full-text patent data on granted patents is to consider the size of the claim set. A larger volume of a claim set of a patent predicts a higher patent value. Counting the number of claims or dependent claims is the best indicator of volume of the claim set. This logic holds both for grants and applications. This makes claim data interesting since the use of applications makes early value recognition feasible.

Combining data on applications and grants also produces an interesting and even better performing value indicator. If there are differences between applications and grants, measured by the number of (dependent) claims, this predicts a higher patent value. The explanatory power of a model based upon differences is considerably higher compared to a model based upon claim size.

When using claim data, one has to take into account that there are structural differences in claim sets. These differences are related to the country in which the patent is filed, the technology field it covers and the year in which it is filed.

18 Discussion

This discussion reflects on this thesis project in two ways. First, the question 'has the research goal been met' is discussed. In this discussion, limitations of and possible flaws in this project are identified, as well as leads for further research. Second, the findings in this thesis project are compared with those in the article of Reitzig because of the important role his 2004-article played in during this project.

18.1 Has the research goal been met?

The goal of this project was to determine if full-text patent data could be used as a patent indicator. Further, it was tested if such an indicator would be available early in patent's lifetime, so that early value recognition would be possible. These two research goals have been met by developing two types of claim indicators: one based upon the claim-set-size and one based upon the differences between applications and grants.

18.1.1 The contribution of a claim indicator

The contribution of the claim-set-size indicator to the already existing patent indicators is limited. When using the number of differences between grants and applications, however the explanatory power increases considerably. Unfortunately, constructing an indicator based upon differences requires more data and more work. As far as the specialists at Philips could tell, the number of claims in grants and applications is not knowledge that is offered clear-and-cut in databases and thus can only be extracted indirectly. Therefore, in practical terms the size of the claim set may still be favorable. As for early recognition, the size of the claim set can predict patent value but its explanatory power is limited and when used together with family size, it diminishes to zero. So, although claims can be used to predict patent value, the actual contribution they can offer is somewhat limited.

18.1.2 Limitations of early value prediction

For early value prediction, there is important limitation of this thesis project. Only patents that were granted were included in the dataset. Therefore, it is difficult to say if the findings also hold for claim sets of applications that were never granted. It is not expected that claim sets of these patents will differ much from the ones that are granted: they will most likely have comparable amounts of claims and claim length. This means that early value recognition based upon claim size will probably fail to take into account the likelihood that a patent will be granted. However, being granted or not is of considerable impact on its value. Also the likelihood of changes cannot be predicted by the size of the claim set.

Therefore, the early recognition model should be enhanced with data that could predict if an application will be granted and if so, whether changes are likely. The number of backward-X-citations has been identified as a possible predictor for this. Unfortunately, with the datasets used in this project it was not able to fully test this assumption. Therefore, future research should determine if x-citations can indeed predict changes of claim sets during the examination procedure or the likelihood that a patent not will be granted. Second, it should examine how this information can be used to enhance an early value recognition indicator based upon claim data.

18.1.3 The claim characteristics that not related to patent value

Another topic of future research can be found in the exploration of the claim characteristics that were identified in this project but that did not relate to patent value. I believe that a more accurate measurement of some of the characteristics can improve their ability to predict patent value. For instance, in 128 cases a first claim was erroneously identified as being both a product and method claim (while a claim can only be or a product or a

method claim). Or, as another example, think of enlarging the set of limiting words so that the concept of 'limitation' is better covered. Perhaps also that some characteristics have more fit when used to explain other concepts than patent value, like for instance the legal strength of the patent.

Although many of the characteristics have not been successfully related to patent value, I still believe they are a contribution. They provide researcher some insights in the implicit rationales that patent attorneys use when writing and reading patent claims and thus can enhance their understanding of what constitutes patent value. <ore on this topic will be said when the Reitzig-article is discussed in the second part of this discussion.

18.1.4 Patent value?

To some extent, the trouble of relating claim data to patent value could be caused by the used value estimations. Patent attorneys at Philips expressed their doubts if inventors could give accurate value estimations. It remains difficult to determine if all inventors were indeed estimating monetary value, as was requested. Perhaps some of them could only make a wild guess, or tried to express other 'value' concepts as for instance the perceived quality of (their!) technology. On several occasions, the Overall Score and the PatVal value gave different results and in general, claim data was better in predicting the Overall Score than PatVal value. Of course, the Overall Score has its own flaws and it is even more unknown to what 'value' concept the Overall Score relates. The point that should be taken here is that when developing patent indicators, it matters how the value of a patent is defined. And once again, this proved to be a difficult nut to crack.

18.1.5 Factors undermining the relation of patent claims with patent value

As was mentioned before, it is conceptually difficult to think through the exact relation between patent claims and patent value. This is due to the endogeneity that patent claims have with patent value, as was mentioned by Reitzig (2004a). Those patents promising to become important might be formulated with much more attention and thus higher quality, and those with better quality are eventually better equipped to become of higher value. However, let us not forget that the most important determinants of patent value are the market conditions. This implicates that erroneous forecasts of patent value could lead to high 'quality' patents having low or zero value, while poorly written claims could end up to be of high value. However, in the latter situation, it is more likely that the patent will be successfully prosecuted. If such events as these would occur, and most likely they do occur, they blur the relation between claim characteristics and patent value. In order to correct for this, future research could investigate to what extent an expectancy of higher value influences the drafting process of patents. Do patent attorneys indeed spend more time on patents with a high expected value? And if more effort is put into patents with higher expected value, is part of it indeed used to make changes and improvements on the claim set during the examination procedure? Answers on these questions could confirm some of the yet untested assumptions that the hypothesis on differences still holds.

18.2 Comparing the findings with Reitzig's article

It is interesting to compare the findings in this project with the article of Reitzig, since the article was a pioneering work on using full-text patent indicators and had a significant impact on this thesis project. Reitzig suggested the use of full-text patent indicators partly because of their early availability, but based his findings on granted claim sets. This project provides evidence for Reitzig's assumption of early value recognition by showing that application claim sets can indeed be used to predict patent value.

One of the biggest problems using full-text data is to *"understand the codification of technology and value-related information by patent attorneys"* (Reitzig, 20004a, p. 943). This project aimed to discover what kind of characteristics in a claim set hint to attorneys that a patent might be of higher-than-average value. In this sense, it

indeed tried to explore the implicit knowledge that attorneys use when analyzing or writing patent claims. Many characteristics were identified. Although in this project most of them were not significantly related with patent value, the characteristics are still interestingly. Their identification and quantification give researchers some grip on the endogeneity of the relation between claims and value, which Reitzig mentioned as the biggest disadvantage of using full-text data.

18.2.1 Comparison of empirical findings

If we compare the empirical findings of Reitzig with this project, we have to take into account that Reitzig used patents from the chemical industry, while this project focuses on patents related to technology fields in which Philips is active. It can be expected that there is a significant difference in the way claims are written in both technology fields. But since the PatVal set used in this thesis project contained a small amount of chemical and pharmaceutical patents and there was a correction for technology fields, still some comparisons can be made. Another difference is that Reitzig uses the likelihood of opposition as dependent variable, while the PatVal set used value estimations of inventors. But the biggest problem is that Reitzig counted independent and dependent claims in a totally different way. He only made this distinction for product claims, while for process or 'application' claims, independent and dependent claims were counted together. This does not seem logical to me, since dependent claims serve a different purpose. Reitzig finds that the number of independent claims is related to the likelihood of opposition, when in fact these independent claims are merely product claims. In this project, however the number of independent claims was not a significant predictor of patent value, even when solely product or method claims were used.

Interestingly, Reitzig mentions the costs of dependent claims as a key reason why inclusion of more dependent claims in a set could indicate higher value. Although fee costs are a rationale why larger claim sets are likely to be of more value, patent attorneys in Philips mentioned such costs are not very influential compared to the overall fee costs (at least not until the recent changes made by the EPO). Thus, Reitzig might have overestimated the importance of costs here.

Finally, both Reitzig's article and this thesis project seem to fail in relating more sophisticated text characteristics to patent value. For instance, he tries to use the number of words in the state of the art or the technical problem as a predictor, but only incidentally he finds significant relationships. Still, I believe that not all hope should be abandoned on digging into sophisticated characteristics. Again, they provide researchers with a view on the tacit knowledge patent attorneys have, which is by itself already a valuable contribution. Reitzig ends his article by saying that "*we must learn to walk before we run".* To end this discussion, I would suggest that this wisdom also applies for this thesis project. Much work still lies ahead in making improvement on early value recognition and the understanding of these sophisticated claim characteristics, but with this thesis, I believe some firms steps were made in the right direction.

19 Literature List

Adams, S. R. (2006) - *"Information Sources in Patents"* - 2nd edition, K*G* Saur, München, 2006.

Albert, M.B., Avery, D., Narin, F., and McAllister, P. (2002) - *"Direct Validation of Citation Counts as Indicators of Industrially Important Patents."* Research Policy, Vol. 20 (1991), pp. 251-259.

Andriessen, D. (2004) - "*Making Sense of Intellectual Capital"*. Elsevier, UK, 2004.

Breitzman, A.F., and Mogee, M.E. (2002) - *"The many applications of patent analysis".* Journal of Information Science, Vol. 28 (3). (2002), pp. 187-205

Criscuolo, P. (2006) - *"The home advantage effect and patent families."* Scientometrics*,* Vol. 66, no. 1 (2006) pp. 23-41

Deng, Yi. (2007) - *"Private value of European patents."* European Economic Review, Volume 51, Issue 7, October 2007, pp. 1785-1812

EPO (2007a) – "*Schedule of fees and expenses of the EPO"*, Supplement to the EPO Official Journal of 11/2007, www.epo.org

EPO (2007b) – *"EPO Guidelines of December 2007"*, part C, chapter 3; www.epo.org

EPO (2008a) – *"Schedule of fees and expenses of the EPO"*, Supplement to the EPO Official Journal of 7/2008, www.epo.org

Ernst, H. (1998) - *"Patent portfolios for strategic R&D planning.*" J Eng Technology Management 1998; 15, pp. 279–308.

Fabry, B., Ernst, H., Langholz, J., Köster, M. (2006) - *"Patent portfolio analysis as a useful tool for identifying R&D and business opportunities – an empirical application in the nutrition and health industry."* World Patent Information 28 (2006), pp. 215–225

Festen-Hoff, K en Rijlaarsdam, A (red.) (2004) - *"Recht voor Ingenieurs"* 5e druk, DUP, Delft 2004

Gambardella, A., Harhoff, D., Verspagen, B. (2006) - *"The value of patents*". In: Paper presented at the NBER Conference, The Economics of Intellectual Properties, Cambridge, MA, July 2006, pp. 1-45

Giuri, P., Mariani, M., Brusoni, S., Crespi, G., Francoz, D., Gambardella, A., Garcia-Fontes,W., Geuna, A., Gonzales, R., Harhoff, D., Hoisl, K., Lebas, C., Luzzi, A., Magazzini, L., Nesta, L., Nomaler, O., Palomeras, N., Patel, P., Romanelli, M., Verspagen, B. (2005) - *"Inventors and invention processes in Europe: evidence from the PatVal-EU survey."* Research Policy 36 (2007), pp. 1107–1127.

Golzio, D. (2006) - *WWWWWHow to Read a Patent!*, European Patent Office, pp. 1-6

Griliches, Z. - *"Patents Statistics as Economic Indicators: A Survey".* Journal of Economic Literature. Vol. 28 (4). (1990), pp. 1661-1707

Grupp, H., Schmoch, U. (1999) - *"Patent statistics in the age of globalisation: new legal procedures, new analytical methods, new economic interpretation".* Research Policy 28 (1999) pp. 377-396

Guellec , D. and Van Pottelsberghe de la Potterie, B. (2000) - *"Applications, grants and the value of patent".* Economics Letters, Volume 69, Issue 1, October 2000, pp. 109-114

Hall, B.H., Jaffe, A., Trajtenberg, M., (2005) - "*Market value and patent citations".* RAND Journal of Economics 35, pp. 16–38.

Harhoff, D., Hoisl, K., (2007) - "*Institutionalized incentives for ingenuity—patent value and the German Employees' Inventions Act*." Research Policy 36, pp. 1143–1162.

Harhoff, D., Scherer, F.M., Vopel, K., (2003) - *"Citations, family size, opposition and the value of patent rights"*. Research Policy 32, pp. 1343–1363.

Harhoff, D. and M.Reitzig (2004) - *"Determinants of Opposition against EPO Patent Grants – The Case of Biotechnology and Pharmaceuticals,"* International Journal of Industrial Organization, 22 (4), pp. 443-480.

Heiden, B. J. (2001) - *"The Microeconomic asset value of a patent - an empirical study of highly valuable Swedish-owned patents.*" Center for Intellectual Property Studies, Chalmers University of Technology, pp. 1-46

Hoeth, L. W. (2007) - "*Kort begrip van het intellectuele eigendomsrecht".* 9th edition, Kluwer BV, Deventer, 2007.

Jackson Knight, H. (1996) - "*Patent Strategy for Researchers and Research Managers"* 1st edition. John Wiley & Sons, Chichester. 1996.

Kürtössy, J. (2004) - *"Innovation indicators derived from patent data"*. Periodica Polytechnica Ser. Soc. Man. Sci. Vol 12, Nr. 1, pp. 91-101.

Lanjouw, J. O. and Schankerman, M. (2001) - *"Characteristics of patent litigation: a window on competition"*, RAND Journal of Economics, vol. 13, pp. 129–51.

Lanjouw, J.O., Schankerman, M., (2004) - *"Patent quality and research productivity: measuring innovation with multiple indicators*." Economic Journal 114, pp. 441–465.

Markman, G.D., Espina, M.I., and Phan, P.H. - (2004) "*Patents as surrogates for inimitable and non-substitutable resources"*. Journal of Management, Vol. 30 (4), pp. 529-544

Meyer, M. (2000) - *"Does science push technology? Patents citing scientific literature*." Research Policy, Volume 29, Issue 3, March 2000, pp. 409-434

Michel, J., Bettels, B. (2001) - *"Patent citation analysis—a closer look at the basic input data from patent search reports*". Scientometrics, 51, pp. 185–201.

Putnam, J. (1996) - *"The Value of International Patent Rights."* Yale University, New Haven.

Reitzig, M. (2003) - *"What determines patent value? Insights from the semiconductor industry"*. Research Policy, Volume 32, pp. 13-26.

Reitzig, M. (2004a) - *"Improving patent valuations for management purposes – validating new indicators by analyzing application rationales"*. Research Policy, Volume 33, pp. 939-957

Reitzig, M. (2004b) - *"What do Patent Indicators really measure - A structural test of novelty and inventive step as determinants of patent profitability.*" Paper presented at the DRUID Summer Conference 2004 on Industrial Dynamics, Innovation and Development, pp. 1 -29

Schankerman, M., Pakes, A. (1986) - *"Estimates of the value of patent rights in European countries during the post-1950 period.*" Economic Journal 96 (384), 1052–1076.

Simmons, E.S., Lambert, N. (1991) - *"Patent statistics – comparing grapes and waterlemons"*. Proceedings of the 1991 International Chemical Conference, pp. 33-78

Tong, X., and Frame, J.D. (1994) - *"Measuring national technological performance with patent claims data*." Research Policy 23 (1994), pp. 133-141

Trajtenberg, M. (1990) - "*A Penny for Your Quotes: Patent Citations and the Value of Innovations"*. The RAND Journal of Economics 21 (1990), pp. 172-187.

USPTO (2009) – *"United States Patent and Trademark Office FY 2009 Fee Schedule"*, last revisited on January 01, 2009, www.uspto.gov, → patents → fees

Van Zeebroeck, N., Van Pottelsberghe de la Potterie, B. (2007) - *"Filing strategies and patent value"*. CEB Working Paper N° 08/016, March 2008, pp. 1-45

Van Zeebroeck, N. (2008) - *"The Puzzle of Patent Indicators"*. CEB Working Paper No. 07/023, May 2008, pp. 1-36

WIPO (2008) - *"World Patent Report: A Statistical Review"*, Edition 2008, p. 14, www.wipo.int

20 Appendix

20.1 Appendix I – An exploration into the most important patent indicators.

20.1.1 Forward Citations

One of the earliest studies relating citations with important patents was done by Reisner at IBM in 1963. However, the famous work of Trajtenberg (1990) put the idea that patent citations could predict patent value on the map. (Breitzman et al., 2002, p. 204) Analogue to the measurement of the impact of scientific publications by references of other publications, it was suggested that the social value of patents can be predicted by the amount of references they receive. These references are called forward citations. (Harhoff et al., 2002, p. 1350; Griliches, 1991, p. 1689)

Forward citations are the most well-established patent indicator. Numerous studies have been done to examine the relationship between forward citations and value and nearly all of them provided significant positive results. (Reitzig, 2004b, p. 7) Among the more well-known studies are Trajtenberg (1990), Hall et al. (2005), and Harhoff et al. (1999). (Giuri et al., 2007, p. 1108) More specifically, it is believed that extra citations lead to higher market value and that self-citations (an applicant citing own work) signal a higher added value compared to normal citations. (Hall et al., 2005, p. 1)

There are several rationales why forward citations are related to value. One idea is that they operationalize the potential of the patent in its market, regardless of the complexity or quality of the invention it protects. Because citations reflect a technological trajectory of (research) investments, more forward citations predict a market for the underlying invention. (Zeebroeck, 2008, p. 5) This matches with the statement of Gambardella et al. (2006) that forward citations reflect visibility. (Gambardella et al., 2006, p. 2) A second claim is that forward citations are good predictors for novelty and inventive step. However, later findings of Reitzig show that forward citations mainly relate to technical importance of a patent. (Reitzig, 2004b, p. 7) Finally, it is mentioned that forward citations given by examiners have reduced the scope of later patent applications, which indicates benefits to society. (Zeebroeck, 2008, p. 5) However, it is not clear what exactly this benefit for society is. The limited scope for later applications only indicate that the cited patent already captured a certain monopolistic position, which cannot be claimed twice. To what extent this can be seen as 'benefit to society' is questionable.

Forward citations appear over time and keep appearing over time: even long after a patent is expired, it can still receive citations. So, ceteris paribus, the older the patent is, the more citations it has received. This makes it hard to compare patents filed in different years. (Hall et al., 2005, p. 5 & p. 13) When comparing between years, an additional bias is caused by the rise in both patent applications and citations per patent over the last decades, creating an additional rise in the amount of citations. (Zeebroeck, 2008, p. 15) It can be expected that especially recent patents profit from this effect by receiving more citations. Another argument not supporting the use of forward citations is that examiners seem to have 'pet patents': patents they cite more frequently than can be reasonably expected. Reasons for this can be rational (the scope of the pet patent is broad and thus frequently delivering prior art) but also irrational (by chance, this patent became well-known to the examiner who therefore uses it frequently). (Simmons, 1991, p. 52-63) Yet another problem that examiners face is language differences, which play a more important role nowadays due to the increasing international orientation of the patent system. Examiners will only cite patents written in languages known to them. Michels et al. (2001) show that all three major patent offices in the world (USPTO, EPO and JPO) are biased: most citations done by examiners from a

certain office are referring to patents processed in the same office. (Michel et al., 2001, p. 197) This is called the 'home-advantage' effect. (Crisuolo, 2006, p. 30)

20.1.2 Backward Citations

A backward citation is the name for a reference (a citation) that a patent gives to relevant prior art. The difference between backward and forward citations is thus made purely by viewpoint: what is a forward citation for patent A coming from patent B, is at the same time a backward citation for patent B towards patent A. Usually backward citations are references to other patents, but scientific literature is occasionally cited also. Applicants in the US are obliged to file all relevant prior art, creating long lists of backward citations, while in the EPO it is mainly the task of the examiner to identify prior art. (Harhoff et al., 2003, p. 1345; Lanjouw et al., 2004, p. 446 and Hall et al., 2005, p. 7) On average, US examiners cite about three times as many patents as their EPO colleagues. (Michel et al., 2001, p. 191)

Citations are written in a search report which aims to seek the base of the patent claims. Thus, the search procedure focuses on retrieving any documentation which can restrict the claim set of an application. In the EPO, references are classified into (mainly) A, X or Y citations. While an 'A' simply means a reference to relevant state of the art without consequences for patentability, an 'X' or 'Y' can pose problems for patentability. Most troublesome is an 'X' citation, meaning that a (set of) claim(s) does not meet the novelty or inventive-step criterion. A 'Y' is less severe, stating that there is no novelty if the examiner combines the content of two relevant documents of the prior art. (Harhoff, 2004, p. 448-449) If sufficient evidence is found for restricting or even rejecting the application, the search process is stopped. Therefore backward citations cannot be considered as an overview of the 'state of the art' of all relevant technology until that moment. (Michel et al., 2001, p. 188)

There are a few rationales for the relation between backward citations and value. Harhoff (2004) suggested that a small scope (and therefore, low value) could be reflected by a large amount of backward citations. Patent attorneys confirmed that broad claims can obtain many backward citations as examiners restrict the scope of such claims. But, even after restriction, they can still be broad. (Harhoff et al., 2004, p. 1350) Reitzig (2004b) puts forward that backward citations relate to "non-technical economic features of the property right" (Reitzig, 2004b, p. 7). In short, the assumption is that they show existing market potential: the more backward citations, the more evolved a market is which means a larger market for the products protected by the patent. To finish off, it is good to stress that some problems troubling forwards citations (the rise in citations, the home-advantage effect) have the same effect on backward citations.

20.1.3 Litigation and / or Opposition

There are several reasons why a lawsuit against a patent is started. In some cases, the complainant has some sort of disadvantage caused by the patent of the defendant. For instance, the scope of the patent prevents the complainant from using a specific technology. In such a case, the complainant can try to limit the scope of the patent by suing the patent owner. Here, the complainant could try to use the argument that the examiner has incorrectly accepted the scope of the patent.

Another reason could be a protective measure. Suppose that company A states that company B is infringing their patent and demands that license fees should be paid. Company B could either accept and pay, or could try to nullify the patent.

Theoretically, whether litigation will take place is dependent on four factors. The first factor is the likelihood of infringement itself. If firms think there is a chance on litigation, they will start a process more soon. The second is the degree of asymmetric information between both parties. In case there is a difference in the perceived outcome

of a trial, the likelihood of litigation increases. This plays especially a role in emerging technologies. Third, if stakes are high (thus if high profits are related to the patent rights), the chances of litigation increase. And fourth, the costs of trial versus settlement have an impact on the probability of litigation. The higher the costs of trial, the more likely settlement will be used. (Lanjouw et al., 2001, p. 132)

Litigation indicates patent value for two reasons. First, there are the high costs that are related with litigation. It is estimated that one trial can cost between 50.000 and 500.000 euro's. (Harhoff et al., 2004, p. 451) Obviously, such costs will only be made if the profits made by the patent are (expected to be) high. The second reason is that litigation indicates strategic interests. Companies filing a lawsuit signal some sort of interested in the technology the patent is protecting and thus expose a bit of their strategy. This can be interesting both for the company sued as well as third parties. (Zeebroeck, 2008, p. 11) Lanjouw adds to this that especially patents forming the base of a technological trajectory face higher likelihood of being sued. (Lanjouw et al., 2001, p. 130)

Opposition, a special case of litigation within the EPO, is been studied by many scholars as a value indicator. It can be compared somewhat with the reexamination procedure within the USPTO although the opposition procedure is used more frequently. (Harhoff et al., 2002, p.1351) In short, the opposition procedure gives anyone the right to file a complaint against a granted patent up to nine months after granting. This can be done for a relatively low price, especially compared to litigation costs in national courts. Opposition is filed with the EPO. It should consist of evidence that the criteria for patentability were not fulfilled and therefore the patent should not have been granted. Outcomes of opposition procedures are binding for all member states. Three outcomes are possible: 1) the patent is upheld without changes, 2) the patent is upheld with changes or 3) the patent is revoked. (Harhoff et al., 2004, p. 449) Around 8% of all grants are opposed, of which 14% is revoked. Ceteris paribus, larger companies tend to have their patents opposed more. (Harhoff et al., 2002, p. 1351)

Opposition of a patent is believed to signal value. A study of Harhoff and Reitzig (2004) has shown that opposition rates correlate with forward citations and family size. They also show that by discriminating between IPC codes, opposition is more frequent in new technological areas. Further, grants that have an extensive record of discussion between the applicant and examiner (shown in e.g. high number of X-type citations or long grant lags) are more opposed than on average. (Harhoff et al., 2004, p. 478)

20.1.4 Renewal Rates

In order to maintain a patent, regularly fees have to be paid to the national patent office. If the fees are not paid, the rights the patent was protecting are no longer upheld. Since the costs of keeping a patent alive are considerable and increase over time, several studies claim that more important patents are kept alive longer, starting with the work of Pakes and Schankerman in 1984. (Gambardella et al., 2006, p. 2; Harhoff et al., 2002, p. 1344) For instance, in the EPO renewal fees range from 400 euro in the third year up to 1350 euro for the tenth year and each subsequent year. (EPO, 2008a) In the USA, the fee schedule of 2009 states that after 3.5 years a renewal fee of 980 US dollars has to be paid, after 7.5 years a fee of 2480 US dollars and after 11.5 years a fee of 4110 US dollars. (USPTO, 2009) Although the fees are remarkably lower in the USA, in both the EPO and the USA the height of renewal fees rises with the age of the patent. The rationale behind this is that a patent lowers the maximum social welfare by the monopoly it creates in favor of the patent owner. The longer the patent owner wishes to keep its monopoly, the more he should be charged for this.

The most obvious problem using renewal rates as a value indicator is timeliness. Van Pottelsberghe de la Potterie and Van Zeebroeck propose a 10-years period counting from filing date as a reasonable threshold to discriminate for importance. (Zeebroeck, 2008, p. 9) For business applications, 10 years is generally far too late.

Another problem with renewal rates is that they appear to be country ánd technology dependent. For Europe, the average patent life is longer in larger economies as Germany and the U.K. This indicates that cross-country comparisons are difficult. Moreover, the age of patents per technology field also differs among countries. Deng (2007) shows that patents related to 'electronics' field have a longer life span than other technology fields if the patents stem from Germany or Japan, while 'electronics' patents originating from the US or the EPO do not have a longer life span. (Deng, 2007, p. 1794-1796)

20.1.5 Family Size

Patents with equal contents filed in various countries form a patent family. The number of countries in which such a family is filed is called the family size. (Lanjouw et al., 2004, p. 447) Although this looks as a robust number, there are actually several family definitions used. More strict definitions (e.g. Inpadoc) use smaller family sizes than wider definitions do (e.g. Derwent). Family size was first introduced by Putnam in 1996. Family size is related to the territorial boundary of protection of a patent. Therefore, it is argued that family size gives an indication of the market size in which the patent is active. (Reitzig, 2004b, p. 7). Another perspective is that since the costs of filing are so high for filing in several countries, only patents with high expected value will be filed in multiple countries. In particular triadic families, which are families filed at least at the EPO, the JPO and the USPTO, are considered to be of high value. (Zeebroeck, 2008, p. 7-8) Family size is seen as a well-established value predictor and has a strong correlation with the maximum age of patent families, which in itself is also a well-established value indicator. (Reitzig, 2004b, p. 7; Harhoff, 2002, p. 1351)

There are also objections against using family size as an indicator for patent value. Guellec et al. (2000) mention that during informal interviews with managers of small high-tech companies, it was suggested that filing in many countries can reflect an immature market. For matured markets, economies of scale are at work. So, only the largest markets in the world have to be protected in order to enjoy world-wide protection. This means that there would be a break-even point after which a larger family size would again indicate lower value. When looking at EP patents, the paper of Guellec et al. identified such a point at a family size of the basic three countries (Germany, France, and England) plus three to four other countries. (Guellec et al., 2000, p. 114) This comes close to the average family size of EPO patent families of 7.9. (Deng, 2007, p. 1791) Another argument is that family size does not take into account the actual markets they cover. Equal family sizes may still greatly differ in the actual number of customers that is reached, merely because of the sheer number of people fluctuates among countries. In addition, different technology fields are related to different geographical settings, further influencing the relation on value. For instance, although patents related to maritime shipping may be of high value, they will certainly not be granted around the world, since some countries do not have industries in this field. (Simmons et al., 1991, p. 36) This is confirmed by the findings of Deng (2007), who shows that family size differs across technology fields. While pharmaceuticals and chemicals score high designation rates, families related to mechanical or electrical inventions score among the lowest family sizes. Further, if inventions are used in multiple technology fields, they tend to have a higher family size. (Deng, 2007, p. 1792)

20.1.6 Filing Strategy

In the literature review different filing routes were discussed: either through a national office, the EPO and or the PCT procedure. Within the EPO and PCT procedures, there are even subdivisions into several specific filing paths, like for instance the accelerated procedure within the EPO. The choice for a certain filing procedure is a tactical choice, which shows some of the company strategy. In case of uncertainty about the profits the patent can deliver, firms can delay the filing procedure. On the other hand, in case expectations are high, the applicant may decide to speed up the process by using a different filing route. (Zeebroeck et al., 2007, p. 12 & Reitzig, 2004a, p. 951)

Therefore, scientists have related the choice of filing route to the (expected) patent value. Based upon filing strategies, policy makers could identify those patents that were more frequently cited, lived longer and were more frequently opposed according to Zeebroeck et al. (Zeebroeck et al., 2007, p. 31) Also certain filing strategies tend to reflect market proximity. (Reitzig, 2004b, p. 7-8) The two most studied procedures are the EPO and PCT procedure, of which the literature findings will be discussed here.

We will start with the EPO procedure. On general level, Deng (2007) shows that patents filed through the EPO route are of significantly more value as those filed through national routes. Most likely, this is because the EPO filing costs are considerably higher. EPO filing is only interesting if one is filing in multiple countries, thus forming a threshold for those patents with low quality (and low expected value). (Deng, 2007, p. 1786 & p. 1799) Within the EPO, there is also an accelerated search and/or examination request which is worth mentioning. If requested, the applicant can obtain his search report or examination more quickly. However, only in 2% of all applications this is requested. The accelerated procedure is also used more frequently for patents with an on-average higher value. (Zeebroeck et al., 2007, p. 12 & 24; Reitzig, 2004a, p. 955)

The PCT procedure was introduced in 1978. Besides it advantage to make worldwide application more feasible, its major advantage for applicants is that they can delay their choice of obtaining a grant from 12 to 30 months. (Zeebroeck et al., 2007, p. 11) The applicant has additional time to explore the potential and expected value of the patent, while at the same time benefiting from the fact that market conditions become clearer as time goes by. (Guellec et al., 2000, p. 112) This extension of both choice and payment is seen as an additional field in which companies can develop patenting strategies. (Grupp et al., 1999, p. 387) The PCT route is associated with higher value and predicts (triadic) family size rather well. Also PCT filings tend to have received more forward citations and have a longer life span. (Zeebroeck et al., 2007, p. 24)

To finish off, a last filing strategy will be discussed which is not examined much in scientific literature: the filing of divisionals. Applicants can split an application into several filings in case the claims of the application lack unity. Hence, divisionals are born each following their own path in the examination procedure. It is believed that divisional filing is used for patenting inventions that are in a very early stage of development, to mislead competitors and examiners or to lengthen the examination process. Therefore, divisionals could be used tactically. (Zeebroeck et al., 2007, p. 14-15)

20.1.7 Applicant and / or Owner Characteristics

Also the aspects of its applicant and / or owner were considered in studies. Do note that the applicant and the owner may differ through patent trading or M&A activities. In other cases, where the applicant is also the owner, it can still be difficult to relate the two. Firms can file patents through their daughter-companies or joint-ventures, thereby causing a less clear relation to the actual owner and making it extremely difficult to process such information automatically. (Simmons et al., 1991, p. 40-44)

Some theories are developed on the relation between applicant and patent value. Zeebroeck et al. (2007) mentions several applicant characteristics that may influence patent value. First, the size of the applicant is expected to influence value. Larger companies are expected to produce more valuable patents. On the other hand, they are less picky on what to file so measurements get blurry. Therefore, even after many studies the relation remains ambiguous. Second, both academic institutions and private firms file patents. The former are expected to produce patents with knowledge related to basic research which could be less applicable in market situations at that moment. Finally the size of the applicants' patent portfolio may indicate value, as it reflects experience with the patent system. But at the same time it can also reflect a lower value, since frequent filing behavior indicates less discrimination towards expected patent value. (Zeebroeck et al., 2007, p. 8; Giuri et al., 2007, p 1122)

There are also some studies in which only ownership is considered to predict patent value. The underlying idea here is that the market value of the owner somehow relates to the value of its intellectual assets, here patents. Stock market value of firms is much used for this. (Hall et al., 2005, p. 9) Also popular is Tobin's Q, which is based upon the famous principle of Noble Prize winner James Tobin. In short, the ratio between market value of an asset and its replacement costs can predict investment decisions. For patents, the imaginary replacement costs for a patent are used as a value prediction for that patent. (Andriessen, D., 2004, appendix A.22, p. 358; Zeebroeck et al., 2007, p. 8)

20.1.8 A Composite Index

Some researchers have combined several of the patent indicators to create a composite index. For instance, Lanjouw and Schankerman (2004) developed an index measuring 'quality to emphasize both the technological and value dimensions of an innovation' by combining the number of claims, forward and backward citations and family size (Lanjouw et al., 2004, p. 443).

The rationale for using a composite index stems from the fact that different indicators are believed to relate to different aspects of patent value. Zeebroeck (2008) states that using a single indicator for value predictions is indeed 'a highly hazardous enterprise'. Indicators can have a high loading on one dimension but a low one on another dimension. As an extreme case, the article shows that only 77 patents out of a set of 370.00 (!) met all five predetermined value thresholds, which were being granted, having a triadic family size, being opposed, surviving opposition and being cited 10 times or more. However, since patent indicators differ in their unit, level of analysis and the moment they become available, it is difficult to combine them into a single index representing economic value (Zeebroeck, 2008, p. 19-20).

20.2 Appendix II – Value Indicators?

Being able to determine the value of a patent is useful in multiple ways. First of all it improves the understanding of innovativeness and the ability to measure it. More specific, it can help to fine-tune the variables in the patent system and turn it into a more stimulating system for innovation. Second, a sound understanding of patent value can enlarge our understanding of the way value indicators as citations or family size operate. Third, if patent value can be better determined, patent portfolios can be optimized and used more strategically (Reitzig, 2003, p. 13-14).

But, what kind of value do these indicators actually measure? For a start, there is the monetary value of a patent. Since patents are used in many ways (protecting inventions, blocking competitors, obtaining licensing revenues, creating standards etc.), there is no standard formula available determining the monetary value for each patent. For instance, monetary value can be determined from the additional revenue a patent owner obtains by its monopolized position. But one could also asses value by determining what a third party would be willing to offer for the patent. Here, the blocking power of the patent plays a more important role. (Heiden, 2001, p. 7) A specific problem related to monetary valuation of patents is the extremely skewed value distribution of patents (Giuri et al., 2008, p. 1121). This skewness means that a small number of patents is of extremely high value while the majority is of little value. This is highly influential to the value for an entire portfolio (Gambardella et al., 2006, p.1).

A second value concept that is important is value for society. It is defined by the welfare a patent will bring to society and relates to the technological importance or impact of the invention protected by it. Addressing a quantitative value to such a concept remains difficult for obvious reasons (Heiden, 2001, p. 8). However, let us keep in mind that the goal of patent statistics that was defined earlier is to highlight those patents potentially more valuable.