LEFIC

Copenhagen Business School
Solbjerg Plads 3
DK-2000 Frederiksberg

Center for Law, Economics
and Financial Institutions at CBS

LEFIC WORKING PAPER 2002-09

Improving Patent Valuation Methods for Management
–
Validating New Indicators by Understanding Patenting Strategies

Markus Reitzig

Copenhagen Business School

www.cbs.dk/LEFIC

Improving Patent Valuation Methods for Management – Validating New Indicators by Understanding Patenting Strategies

Markus Reitzig[1]

The Copenhagen Business School

May 2002

Abstract:

Indicator-based methods that enable inexpensive evaluations of patent rights appear to have great potential as management tools. However, as of today these methods still require refinement to satisfy companies' applied needs. This paper analyzes the validity of so-far untested indicators of patent value to enhance the quality of patent assessments using indicators. Following an overview of the state of the art, the article expands the theory by eliciting patent attorneys' strategies to maximize profits from protecting intellectual property. Inspirations for the computation of new value indicators are gathered. Then, based on a newly compiled data set consisting of 813 EP patents, the probability of an opposition against a patent is modeled by established and new value indicators. The untested indicators draw from publicly available procedural information as well as full-text documents. The results show that accelerated examination requests and qualified word counts are correlated with the opposition decision and enhance the quality of existing valuation methods.

Keywords:

Patents, valuation methods, value indicators, strategic patenting, chemical industry

---

[1] Markus Reitzig, Department of Industrial Economics and Strategy, The Copenhagen Business School, Solbjergvej 3, 2000 Copenhagen-Frederiksberg, Denmark.
Email: reitzig@cbs.dk.

Research performed while author was a working with the University of Munich and as a visiting scholar at the Walter A. Haas School of Business, Business and Public Policy Group, CA 94720-1930 Berkeley/USA. The research was funded by the German Academic Exchange Service (DAAD) through grant no. D/01/04933.

Introduction

"Most investment banks have teams of accountants, tax advisers, management consultants, and regulatory affair experts to structure their deals to a company's greatest advantage. But one would be hard-pressed to find a major investment bank that employs even one individual with experience in evaluating patent portfolios. [...] as matters stand now, 'due diligence' regarding patent assets is usually more myth than reality." (Rivette and Kline, 2000)

This critique by Harvard Business Review authors Rivette and Kline is harsh. Existing services offered by investment banking houses to value intellectual property (IP) are given little credit; serious doubts are especially uttered concerning the practitioners' expertise and competence in evaluating patent *portfolios*. At the same time, the authors foresee a rising importance of IP assets in corporate business strategy.

From my knowledge of the field I agree with the authors' remarks. Looking at the scientific literature, however, I find it hard to put the blame on the practitioners. As a matter of fact, despite the diversity of articles from Industrial Organization (IO) or legal scholars on value related issues of intellectual property rights, there is a lack of scientific papers that restructure the knowledge on the evaluation of patent rights from a corporation's perspective. Building on earlier works by Pakes (1986) and Harhoff et al. (Harhoff D, Scherer F, Vopel K. 1999. Citations, Familiy Size, Opposition and the Value of Patent Rights. Ludwig-Maximilians-Universität München, Harvard University, ZEW Mannheim, Munich/Boston/Mannheim), Reitzig (2002) lays out that there exist several approaches which can be pursued by companies to value their intellectual assets. As it turns out, valuation approaches using patent indicators seem especially convenient for the assessment of patent portfolios comprising a large number of intellectual property rights. Here, indicators drawing from publicly available patent data banks are computed for individual patents and are fed into a valuation algorithm that yields the patent portfolio value as the cumulative value of the individual patents. Those patent indicators can usually be computed at little cost per patent. However, both from a theoretical and applied standpoint the indicators need to be valid correlates of patent value. Furthermore, the indicators used should be available early in a patent's life to allow for evaluations of young patents that may be particularly interesting for the company's future performance.[2] As the prediction quality of the portfolio's value normally increases with the number of valid patent correlates used in the estimation (if they are not collinear), there exists a vital interest in validating as many indicators as possible.

The scientific challenge at this point therefore lies with the validation of further patent indicators that draw from publicly available information and are available early in the lifetime of the patent. The task is especially aggravated by the complex interdependencies between a patent's economic value, the latent determinants of this value, and observable information resulting finally from legal actions that can be used to compute indicators.

This paper addresses this problem in two steps. First, I provide a theoretical framework by laying out the state of the art and then expanding the existing theory of measuring patent value with indicators. Then, empirical results from a large scale study in the chemical industry are presented. In more detail, the remainder of the paper is structured as follows: Part two of the paper addresses theoretical issues. The third section of the paper describes the research design. Here hypotheses concerning the correlation between a patent's economic value and the indicators are derived. Part four of the paper presents empirical results. The paper concludes with a summary providing an outlook on future research.

[2] Note that from an *ex-ante* point of view (filing date of the patent) explanatory variables are already valid patent value indicators if they are correlated with the *anticipated* value of the patent. This paper considers indicators as

Theoretical Framework

The following section is split in three paragraphs. First, it is worth reviewing briefly a definition of patent value that is suited for companies which find themselves in a competitive environment. Secondly, an overview of the state of the art on value indicators of patents is presented before I finally move on to a third subsection in which I try to open up the 'black-box' of patent attorneys' work. These last results are based on interviews with nine senior patent experts from several law firms and the European Patent Office (EPO). The description of the patent attorneys' strategies to maximize profits from protecting IP inspire the computation of new value indicators that use observable information of two kinds: Procedural information and full-text information from patent drafts. The latter type of information in particular has rarely been used to compile indicators of patent value to date.

A definition of patent value

The value of individual intellectual assets is rarely observable. This may be due to the fact that almost no marketplace exists where single patents are dealt. The great idiosyncracy of a patent's value for different potential owners associated with information asymmetries between potential sellers and buyers, and the dependence of a single patent on its surrounding portfolio of IP assets may be seen as the reasons for this phenomenon.

Thus, to determine the value of an individual patent, inductive approaches must be chosen and a definition for the latent construct 'patent value' is needed. Harhoff et al. (1999) show in a formalized fashion that for a corporation involved in technological competition, the value of a patent is best defined as its asset value. This definition covers the majority of the empirically

---

'valid' if they either correlate with the patent's value from an *ex-post* or from an *ex-ante* perspective.

relevant scenarios. To determine a patent's value, it is therefore necessary to consider its (observable) effects on *prices*, *costs*, and *sold quantities* of patent-protected products by the owner and its simultaneous (unobservable or counterfactual) effects on the proprietor's competitors. As Reitzig (Reitzig M. 2001b. Evaluating Patent Portfolios - Using Indicators for Technology Management Purposes. Ludwig-Maximilians-University Munich: Munich Germany.) shows in a survey of the theoretical literature, counterfactual effects should become assessable when quantifying the following patent's latent value determinants: *state of the art (of existing technology)*, *novelty*, *inventive activity*, *breadth*, *difficulty* to *invent around*, *disclosure*, and *dependence on complementary assets*. Thus, when speaking of indicators of patent value, they can be theoretically valid correlates of a patent's value in two fundamentally different cases. Either they show a direct correlation with observable *prices*, *costs*, or *sold quantities* of the patent protected product, or they operationalize latent determinants of patent value such as *novelty*, *inventive activity*, *breadth*, *difficulty* to *invent around*, *disclosure*, and *dependence on complementary assets*. Figure 1 illustrates the interdependencies between patent indicators (examples are forward citations, backward citations, and family size), observable economic quantities, and latent determinants of patent value.

Insert Figure 1 about here

As of today, however, little empirical evidence exists on the complex interaction between indicators, determinants, and prices, costs, and quantities of protected products sold. Moreover, even from a theoretical point, the complex information hidden in the patent data is still a 'black box' to many economists. In the context of this paper, it seems particularly puzzling that indicators can refer to different determinants of patent value at the same time. Claims, for example, have been related to the *breadth* of a patent. At the same time, they also reveal information about its *inventive activity (non-obviousness)*. As *breadth* and *inventive activity* may affect the economic value in different ways, however, these ambiguities pose problems on the interpretation of the coefficient of the claims indicator on value. The same problem holds true for several other indicators, especially for those indicators that use highly patent specific-information. This paper therefore tries to contribute to a better understanding of interactions and interdependencies between patent value, value determinants, and indicators by analyzing patent attorneys' decisions during the patent application procedure. By doing so, the paper also inspires the compilation of value indicators using so-far unused patent information. Before this analysis is undertaken, however, the existing state of the art on the assessment of patents using indicators is briefly summarized in the next section.

Known indicators of patent value – an overview

Until today, a variety of variables have been tested as indicators of patent value in empirical surveys. Looking at 23 empirical studies related to patent indicators and value, Reitzig (2001b) analyzes the appropriateness of the 13 best-known indicator variables for business purposes. In the following, only the results are presented (for a detailed description, see Reitzig (2002), chapter 4). Table 1 summarizes known patent indicators and their advantages and limitations for business purposes. The three columns in Table 1 each refer to one of the evaluation criteria for patent indicators laid out in the introduction to this paper. Column A reports on the validity of the indicator variable. Column B shows the point in time at which the information to compute the indicator becomes accessible. Time is measured in months starting from the filing date of the patent.[3] Finally column C reveals whether the information is available

[3] Note that the information on time is only valid for DE or EP patents.

electronically or has to be collected manually. All indicators draw from publicly available information.

Insert Table 1 about here

Column A itself is subdivided into two subcolumns that regard the theoretical plausibility and the existing empirical evidence for the validity of the indicator in separate ways. It turns out that forward citations, family size, and the ownership variable show the highest degree of theoretical and empirical validation. However market value also seems to be a good indicator for a company's intellectual property assets.[4] Forward citations had been introduced by Trajtenberg (1990) and had been validated as indicators of patent value in numerous subsequent surveys, e.g. by Albert et al. (1991), Harhoff et al. (1999), Lanjouw and Schankerman (2000), and Harhoff and Reitzig (2000). The rationale standing behind the use of this information as a measure for patent value is that the economic importance of a certain patent should be correlated with the frequency at which it gets cited as relevant state of the art for further developments. Family size was introduced as an indicator by Putnam (1996) and again re-validated by Lanjouw and Schankerman (2000), Harhoff and Reitzig (2000), and Guellec and van Pottelsberghe de la Potterie (2000).[5] The idea standing behind the use of this measure is that patent owners signal their willingness to incur increased costs for international patent protection which is associated with increased returns from patent protection, too. On the one hand, the patent owners are apparently willing to incur additional fixed application costs. On the other hand, they signal that they are willing to run an increased risk of costly legal arguments. The correlation between

[4] Note that the 'market value' indicator differs from the other indicators in three respects. First, the market value of a company only allows to serve as an indicator of the aggregate value of intellectual property assets of the company. Besides, almost all empirical studies on the correlation between market value and the number of patents report on a lag structure which has to be taken into account. Finally, 'market value' is information that does not draw from publicly available patent databanks.

market value and patents had been examined by Griliches (1981), Conolly et al. (1986), Conolly and Hirschey (1988), Cockburn and Griliches (1988), Megna and Klock (1993), and Hall et al. (2000). All the studies mentioned above differ with respect to the quality of the research design, the sample sizes, and the kinds of patents (US, EP, DE). They do, however, have a common feature in that they all validate indicators which are linked to patent value by rationales that speak to rather general economic considerations which do not particularly involve in-depth knowledge of institutional details of the patent system. The concept of using citation measures was well known from other disciplines of social science. The fact that ownership affects value is a classical IO consideration. Thus, these indicators may be seen as 'first generation' indicators of patent value. By saying so, no depreciation whatsoever is expressed. On the contrary, the indicators seem reliable and helpful for the evaluation of patents.

In more recent times, other observable information from patent databanks was taken to compile further proxy variables of patent value. In his study, Lerner (1994) successfully linked the market value of 535 biotech companies to the number of patents and the average number of 4 digit International Patent Classifications (IPC) of the company's patents. His goal was to operationalize the 'breadth' or the 'scope' of a patent. Unfortunately, the 'scope' variable turns out to be an insignificant regressor in most of the subsequent surveys. Guellec and van Pottelsberghe de la Potterie (2000) and Harhoff and Reitzig (2000) computed further indicators speaking to patent-specific economic considerations, such as referring to the filing strategy or the legal contents of backward citations (i.e., how many patents in the relevant technical field did already exist before and how similar are they to the patent that is to be evaluated). Obviously, as of today there exists less empirical evidence for these 'second generation' indicators that use patent-specific procedural information and link it to patent value or patent value correlates. Still,

---

5 See Guellec, D. and B. van Pottelsberghe de la Potterie (2000). Analysing Patent Grants. Free University: Brussels Belgium.

the indicators are appealing as they take on the patent-specific knowledge and use it for the computation of value proxies.

Up until now, however, very few researchers have exploited the last resource of information available on patents (i.e. the patent full-text documents themselves). Both, 'first and second generation' indicators make use of 'first page' information stored in databanks. To patent attorneys, this seems strange to some extent, since most of the information on a protected technology and its anticipated economic value is conveyed in the patent draft itself. But then again, special knowledge is required to decipher the relevant information which is codified in the patent document in a very special kind of way. Tong and Frame (1992) were the first to use information from patent documents and make an attempt to compute what I will call the 'third generation' indicators. They correlated the number of claims in a patent draft to several macroeconomic indices of a nation's technological performance. Most recently, Lanjouw and Schankerman (2000) utilize the information on claims to model the probability of challenge and validity suits for a sample of US patents. The number of claims has been regarded as a possible operationalization of a patent's 'breadth'. Third generation indicators (i.e. any indicators compiled from the patent full-text itself) seem to have one major advantage and one major disadvantage over other indicators. They are attractive since they are available early in time (directly after the publication of the patent) and since they show a strong theoretical foundation. Their disadvantage lies in their endogeneity, meaning that the patent document is drafted by the proprietor (or his attorney) who therefore has the opportunity to infer on the value of his patent by the mode of drafting the document.

Still, when thinking of ways to develop new value indicators, the greatest potential lies with second and third generation variables. The challenge here is to understand the codification of technology and value-related information by patent attorneys in such detail that compilations of new indicators show a maximum of theoretical foundation, and a minimum of ambiguity and endogeneity. The following section therefore sketches the strategic considerations followed by patent attorneys during the filing process and opens up the black box of their codification.

Expanding the theory – opening the "black box" of patent attorneys' work[6]

To gain an understanding of the meaning and interdependencies of observable patent information and link it to patent value, interviews were conducted with nine senior experts from patent law firms, a corporation's patent department, and the European Patent Office (EPO). As it turned out, the core of the patent attorneys' work is to maximize profits from legal protection for a given invention. Economically speaking, the patent attorneys' work comes closest to a decision-making problem under uncertainty. I will therefore first outline the decision problem in an abstract way. Then I show how exogenous and endogenous variables (from the standpoint of the patent attorney) enter the attorney's rationale. I will focus mainly on the *state of the art*, the *inventive activity*, and the *breadth* of the patent. Since those decision variables are latent variables, I will finally outline how the attorney's rationale translates into observable action. Here, I will focus on the draft of the patent application and briefly mention two procedural steps that have not yet been described in the literature.

The following descriptions of the decision-making problem the attorney faces refer to the European patent system. Thus, some procedural details cannot be directly transferred to the US system. The basic material trade-offs, however, also hold true for US patents.

Patent protection in Europe can be achieved in three ways. Either the applicant chooses separate national filings in the countries in which he/she seeks protection or he/she decides to file a central European application according to the European Patent Convention (EPC) leading to a

[6] An extended version of this paragraph appeared as a separate German publication by Harhoff and Reitzig (2001).

European (EP) patent. A third possibility is to use a global priority (PCT) application and subsequently decide for one of the two ways described above. The modes differ with respect to fixed and variable costs. As a rule of thumb it may be stated that the fixed costs of filing increase going from the national, via the European to the global application mode. At the same time, variable costs for additional designated states of protection decrease in the same order. As the data set in this paper is based on patents filed exclusively via the EPC or PCT, the description of the decision-making process is limited to EP patents only. The 'life' of a patent in Europe may take several paths. After its grant it can be centrally legally attacked in a so-called opposition procedure within nine months. Third parties gain the chance to diminish or completely destroy the patent's validity for its entire territory of legal effect. The territory of legal effect is chosen by the patent holder. He designates the countries for which he seeks protection and incurs variable costs for each country. The EPO decides on the opposition filed and either upholds, amends, or revokes the patent. Appeals against decisions on the opposition plea by the EPO can be filed from either side, the patent holder and the opposing party. Figure 2 shows the legal 'life-tree' of an EP patent.

Insert Figure 2 about here

Using the tree in Figure 2, the patent attorneys' work can now be described in an abstract way. (Anticipated) Profits can be assigned to all the outcomes of the tree. Probabilities can be assigned to the occurrence of the different legal scenarios (not illustrated in Figure 2). The patent's value is then the sum over the expected profits (i.e. profits times probability of scenario) in all possible scenarios. It is the job of the attorneys (in cooperation with technology managers) to influence profits in discrete scenarios and probabilities of different scenarios becoming true in such a way that the overall expected profits are maximized.

According to the experts, the *state of the art*, the *inventive activity,* and the *market size* underlying the protected invention are the most important exogenous parameters in the maximization process. Besides, the industry often dictates whether the patent can be used as an *exclusion right* in the traditional sense or whether it may rather serve as a *bargaining chip* in technology negotiations with other companies (see Rahn, 1994, and Hall and Ham, 2001). The most important set-screws to be influenced by the patent attorney on the other hand are *breadth*, *disclosure*, and the *mode of filing*. What makes the maximization process complex is that the endogenous variables influence the patent's overall expected value in opposite ways through the probabilities and the static profits. In fact, trading off between the different effects of the endogenous variables is therefore a crucial part of the attorneys' work as will become clear from the following.

At the first meeting between patent attorney and inventor, the expected net profits from protecting the invention are assessed, basing the estimation on the exogenous parameters mentioned above. The estimations are very qualitative, but this is how the attorneys value the exogenous variables:

- Little *state of the art* hints at maximum at a 'latent' market where benefits from patenting can be expected in the future.
- 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.
- If *inventive activity* is small and there is little *state of the art*, expected profits are small.
- If *inventive activity* 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 raises, too, decreasing the overall expected profits from patenting. Expected profits may range from medium to high.

- If *inventive activity* 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.

Given the exogenous variables, the patent attorney can maximize profits by adjusting the endogenous variables with respect to the situation he/she is facing. He/she will extend the *breadth* to its maximum for patents showing a high *inventive activity* and possibly high profits. By doing so he/she maximizes the profits for each scenario in Figure 2. He/she may well increase the probability of a legal attack at the same time, but the probability of losing in the opposition case is small. The fixed costs for the opposition are outweighed by the increase in the profits. Conversely, the attorney will reduce the *breadth* for patents with a decreasing *inventive activity*. The higher the possible profits from a valid patent the more he/she will reduce the *breadth* given the same *inventive activity* since he/she does not want to lose the patent in a legal dispute. The considerations are similar though slightly different for bargaining chip patents. Here, legal disputes are the exception and the attorneys will only make sure that the application 'survives' the granting procedure.

Until this point, consideration was only given to latent variables that drive the rationale of patent attorneys. The attorneys' considerations, however, manifest themselves in the patent draft. Thus by looking at the patent draft, it should be possible to gain hints at the anticipated value of the patent by the attorneys. The interviews reveal that this task may in practice be aggravated by the fact that different patent attorneys have individual *modes of drafting* and that considerable noise should be expected when pursuing a patent text analysis. Still, in principle the following passages in the patent draft should reveal the information of interest:

- The *state of the art* is described in the first section of the patent.

- The degree of *inventive activity* is reflected in the description of the technical problem. The technical problem is normally presented following the description of the *state of the art*. Its solution is presented in the *disclosure* of the patent, and summarized in the
- Claims' section at the end of the patent. Claims also refer to the *inventive activity* behind the patent. At the same time, the *breadth* of the patent should be reflected in the claims. In the chemical industry, especially the number of independent product claims should be an indicator of patent *breadth*.
- Dependent product claims, process- and application claims also add to the *breadth* of the patent. At the same time they operationalize what patent attorneys call fall-back options for legal disputes. Their number should rise with an increasing risk of legal attack (falling *inventive activity*, increasing profits in scenarios).
- Finally, technical advantages and preferred technical solutions in the *disclosure* should also serve as hidden fall-back options. On the other hand, they often demonstrate that inventor and attorney already have an application of the invention on their mind, pointing at an existing market.

Figure 3 illustrates the decision making during the filing process and links it to the observable information described above.

Insert Figure 3 about here

As mentioned above, the decision-making process of the attorneys takes place under uncertainty. Thus, before drafting the patent application, attorneys will try to gather as much information about the underlying *state of the art* and the market size as they can. The information will *ceteris paribus* enhance their ability to assess the patent's *novelty* and *inventive activity* and hence its economic value. A way to gather information more quickly than usual is to request an accelerated search report on the state of the art from the EPO. A way to "buy" decision time is to file the patent through the PCT.[7] Once attorneys decide that protection is valuable and should be acquired as soon as possible, they can accelerate the granting procedure in the European system by requesting an accelerated examination. On the global level, they can accelerate protection by applying through the so-called chapter II of the PCT.[8]

The Empirical Research Design

To validate new indicators of patent value, this paper attempts to link patent value to observable procedural information and to the design of certain text passages in the patent draft in a large-scale empirical study. As valuations of patents are very hard to get, a patent value correlate is chosen as the dependent variable in the regressions, namely the likelihood of an opposition against the patent. In the following I will briefly sketch why the approach seems plausible in general but I will also point at the interpretation problems of the regression results that occur from the chosen design.

Opposition and patent value

Extending the model by Lanjouw and Lerner (1997), Harhoff and Reitzig (2000) can show that the condition for the occurrence of an opposition is given by formula 1 (see Harhoff and Reitzig (2000) for the complete derivation. In the following only the main findings are recalled):

[7] Further details follow in the interpretation of the multivariate statistical results.
[8] See footnote no. 6.

$$P = 0 \quad if \quad ja \geq \frac{jW - (L + l) + S}{w} \qquad (1)$$
$$P = 1 \quad o.w.$$

In formula 1, ja corresponds to the value of the valid patent for its owner and j is the benefits of a successful opposition for the opponent. W is the anticipated probability by the opponent of winning the opposition, w is the patent holder's anticipated probability of losing the patent. L and l refer to the litigation costs for both parties, and S are the settlement costs. Formula 1 illustrates that the probability of an opposition is correlated with the value of the valid patent for the patent owner, ja. This observation supports the research design chosen in this study. At the same time, however, formula 1 also shows that the likelihood of an opposition depends on probabilities of the opposition outcome as anticipated by the opposing parties. In fact, if settlement costs exceed litigation costs by large, the settlement option becomes negligible and the likelihood of an opposition is described by formula 2.

$$P = 0 \; if \quad j \geq \frac{L}{W} \qquad (2)$$
$$P = 1 \; o.w.$$

Thus, when interpreting regression results of the likelihood of an opposition on indicators, two things should be kept in mind. At first, the likelihood of an opposition is driven by the profits j of the *opponent* in the case of a successful opposition. Those should be highly correlated though not necessarily identical to the value of the valid patent for the *owner*, ja. In a simple one-product world where the patent protects a single product and there are only two players, j would be the duopoly profits of the *opponent* whereas ja would be the monopoly profits of the patent *owner*. To faciliate the following descriptions, I will refer to the opponent's benefits from a successful opposition as the *patent's value*. Assuming that j and $j\alpha$ are similar, the patent's absolute value will be similar for both, the patent owner and the opponent. Secondly, proxy variables may well refer to both the value of a successful opposition for the opponent as such, and the anticipated probability of the outcome of the opposition procedure.

<u>Data collection and computation of indicators</u>

The only available source of patent full-texts in machine-readable format for the time being is the EUROPATFULL© databank maintained by the Fachinformationszentrum Karlsruhe/Germany. At the day of the data collection, the full-text patent data were only available for EP patents granted between 1992 and 2000. Given the average time of around 4.3 years for granting a patent at the EPO, I chose patent filings for the years 1992 - 1994 for the study. I decided to focus on patent filings from the chemical industry so that the patent rights would be exclusion rights rather than bargaining chips (see e.g. Cohen et al., 2000). The sampling was based on a four digit IPC classification for industries as proposed by Schmoch and Kirsch (1994)[9]. The patents come from six different chemical branches: organic fine chemicals (37%), polymer chemistry (38%), pharmaceuticals (5%), biotechnology (3%), agricultural chemistry (1%), and petroleum chemistry (16%). As the computation of indicators referring to the wording of the patent draft should originally be carried out using a text scanning software for German language, I also decided to look only at patents who would have a German, Swiss, or Austrian inventor.[10] Out of a sample of 2570 remaining patents I chose 1000 patents almost randomly, the only alteration being to ensure that the opposition rate would be on the same order of magnitude

[9] Appendix A shows the definition of industrial sectors using four digit IPC classes as proposed by Schmoch and Kirsch (1994).
[10] The software searches for German keywords in the patent draft and can therefore only be applied to German documents for the time being. First results showed, however, that for the time being the software still yields significantly different results from a manual compilation which is why the indicators were computed manually for this study once more.

as the long-term average rate of 8.1% for all EP patents. Finally, out of the 1000 documents 813 appeared to be non-truncated and complete with respect to all different sections in the draft and were used for the analysis.

Using three further databanks[11], indicators drawing from procedural patent information were computed. Among the indicators listed in Table 2 are 10 indicators (the first 10 in the table) that have been used in earlier studies. The last two indicators refer to the patentee's options to accelerate the production of the search report by the EPO or to accelerate the granting procedure at the EPO which have not been tested before. Table 2 lists the variables in the first column, reports on their computation algorithm in the second column, and states the data source in column four.

Insert Table 2 about here

Contrasting Tables 1 and 2, it is conspicuous that forward citations are not computed in this study. The rationale behind this is very straight forward. The obvious disadvantage of forward citations as value indicators is their late availability in time. The goal of the paper, however, is to present indicators that serve as proxy variables for patent value at an early stage of the patents' life.

Aside from the indicators utilizing procedural patent information, I computed variables that directly draw from the full-text of the patent draft. The indicators are described in Table 3. The first column of Table 3 names the indicator, the second one briefly recalls the link between the indicator and the economic value of a patent.

[11] The following three data banks were used in addition to EUROPATFULL©: REFI, ELPAC, and EPASYS. The first two are all commercially available. Access to the last one was provided by the European Patent Office in Munich/Germany.

Insert Table 3 about here

<u>Derivation of hypotheses</u>

Column three of Table 2 and column five of Table 3 show the expected signs of the variable coefficients when tested as correlates of the likelihood of an opposition. Whereas correlations between procedural indicators and the likelihood of an opposition should be primarily mediated via the patent’s value, there seems to be a more complex relation between the text indicators and the likelihood of the opposition. Columns three and four in Table 4 therefore distinguish between the expected sign of the correlation between the profits from the protected invention and the text indicators, and the anticipated probability of the opponent to win in an opposition and the text indicators respectively. Column five of Table 3 then provides a very preliminary expectation of the aggregate effect that text indicators should have on the likelihood of an opposition.

Due to possibly counteracting effects associated with the text indicators and resulting ambiguities concerning specific expectations in the empirical study, I chose to test three hypotheses of reduced information contents in this paper. The hypotheses tested in this paper are the following:

- H1: There is a correlation between procedural indicators and the likelihood of an opposition.
- H2: There is a correlation between text indicators and the likelihood of an opposition.
- H3: There is a correlation between a set of procedural and text indicators and the likelihood of an opposition that is stronger than the correlations described in H1 and H2.

Empirical Results and Discussion

In the following, the empirical results are presented and discussed. The section on descriptive statistics will present the data and briefly report on the differences in the means for several explanatory variables in the group of the opposed patents and the remaining group. This leads to the description of the multivariate analysis in which I regress the likelihood of an opposition on the indicators. Following the hypotheses presented above, I will present separate regressions, based on indicators using procedural information, text indicators, and a combination of both separately. The aim is to discover whether the new text indicators are substitutes or complements to the established and new indicators drawing from procedural data.

Descriptive Statistics

Table 4 summarizes the data as they were used in the study. The upper part of the table refers to procedural explanatory variables, the lower part shows the means for the text indicators.

Insert Table 4 about here

As can be seen from Table 4, about 13% of the patents in the sample were opposed. Thus, the opposition rate in the sample is slightly higher than the long-term average of 8.1% for the industry. Patents in the sample might therefore be a little more valuable than on average for this industry. Moving on to the other procedural information, some peculiarities can be observed. Whereas the number of inventors and applicants seem very plausible comparing them to earlier studies, the PCT application ratios appear to be quite low. In fact, further cross checks of the data with the official bulletin issued by the EPO (EPO, 1998) confirm that the low percentage can be attributed to the selection criteria of the sample. With respect to the requests for accelerated search or examination, the means in the sample again correspond to the long-term average value for EP patents across industries and seem therefore plausible. Going further down in Table 4, some observations seem noteworthy when looking at the text indicator variables. At first, all of the explanatory variables show remarkable variation which is intuitively positive. With respect to the number of independent claims, the mean value of 0.64 deserves some explanation. In order to distinguish between independent product claims and other independent claims (process or application claims), I counted product claims separately. Thus, the number of independent claims only refers to product claims, so does the number of dependent claims. Process and application claims were counted separately, but here no distinction between dependent and independent claims was made.

Having checked on the plausibility of the data and their appropriateness for the study, a more focused look on the variables' means with respect to the underlying research question is taken in Table 5. Table 5 shows the means for the explanatory variables computed separately within two different categories, namely patents that received an opposition, and other patents.

Insert Table 5 about here

Table 5 shows that for five of the explanatory variables significant differences in the variables' means exist in the two chosen categories, namely the family size, the number of inventors, the PCT II indicator, the indicator for an accelerated examination request, and the share of A-documents among the patent backward citations.

The difference in the means for the family size shows the expected sign. Increasing variable costs for additional designation states should correspond to a higher (anticipated) value of the patent and hence the opposition rate should rise. Coming to the number of inventors, one would argue that the technical complexity of an invention is likely to increase with the number of people involved in its discovery. Looking at the differences in the PCT II means, theoretical expectations are confirmed by the data. A straight-forward explanation is the following: Higher fixed costs for patent protection should be correlated with higher anticipated revenues. The opposition rate should therefore be higher for patents filed through PCT II. In fact, the following section on the multivariate analysis will present more subtle explanation patterns that are especially needed when discussing the insignificance of the PCT I indicator. Again, a simple explanation for the increased mean of accelerated examination requests in the opposition category speaks to the *cost commitments* made by the patentee. The patentee incurs the risk of sunk costs when filing the accelerated examination request as he pays all fees at a point where the grant of the patent is not yet guaranteed. Speaking with Harhoff and Reitzig (2000) I would also have expected the share of A-documents among the backward citations to be lower in the opposition category. A higher share of A documents hints at ‘non-dangerous’ *state of the art* (i.e. *state of the art* which does not threaten the patent’s legal validity).

Coming finally to the text indicators, again only some preliminary considerations shall be mentioned that are extended on shortly. Two fundamental explanations can be applied to the results of Table 6. Either the word counts do not serve as proper operationalizations of the latent value determinants of a patent and do neither refer to the anticipated value of the patent otherwise, or the correlations between the text indicators and the likelihood of an opposition are distracted by other effects in the data. Here again two explanation patterns seem most plausible. Either differences across the various chemical branches in the sample overlie the correlation proposed in H2, or individual modes of *drafting* patent documents by different patent attorneys lead to a systematic perturbation in the data. In the multivariate analysis, I try to ‘filter out’ both possible effects. I control for the chemical branches and correct for heteroscedasticity in the data that may occur from the individual filing modes of the different patent attorneys.

Multivariate Analysis

Seven different regressions of the likelihood of an opposition on value indicators are shown in this section. Table 6 presents three estimations of the likelihood of an opposition using a simple probit model based on indicators that draw from procedural patent information. Besides, dummy variables for the separate chemical branches enter to ensure that industry effects are not attributed to explanatory power of the indicators.

Insert Table 6 about here

Specification 6A is significant at the 0.1% level. Besides, five variables are individually significant, namely the family size, the number of inventors, the PCT II indicator, the indicator for an accelerated examination request, backward citations to the patent literature, and the dummy variable for organic fine chemistry. Thus, most of the bivariate results presented in Table 5 are reconfirmed in a multivariate context controlling for differences in the chemical branches. The insignificant coefficients in Table 6 shall be given only little consideration. As laid out in the survey of known indicators, I doubt that the 'scope' variable is a valid measure of the patent's *breadth* and thus the result is not too surprising. The number of applicants might be insignificant due to the sample composition. As a matter of fact, multiple ownership was observed in only 15 out of 813 cases. On the other hand it is rather surprising that A- and X-citation classifications turn out to be insignificant in this study. Turning to the PCT indicators, it seems interesting that a filing according to chapter II of the PCT is highly correlated with the likelihood of an opposition whereas the PCT I indicator is insignificant. Therefore, a simple rationale relating the PCT indicators to patent value speaking only to increased fixed costs seems unconvincing. In deed, the argument explaining the empirical results is a little more complicated. Applicants filing patents through the PCT may choose the option for two different reasons that are properly to each other. Either they are very uncertain about the economic success of the patent's underlying invention and they choose the option to "buy" additional decision time as will be described in the sentence after the next one, or, on the contrary, the economic success of the patent's underlying invention is free of doubt already at the date of filing and the option is used to seek global protection as fast as possible. The fact the PCT can be used in these two opposite ways can only be explained when looking at the institutional details of patent law. PCT filings include a search of the *state of the art* by so-called International Search Authorities. The search report is produced within 18 months after the day of filing and then published. 19 months after filing the PCT application the patentee may either decide to drop his application if he/she thinks it is not profitable to seek protection. Alternatively, he/she may initiate the actual examination process and take a costly decision to pursue the filing until the end. He/she may finally also vote for a procedural way that is offered by chapter II of the PCT. This last procedural way offers the applicant an opportunity to request a so-called preliminary international examination and gives him/her another 11 months until he/she finally has to decide whether he/she enters the costly application process until the end. Thus, the PCT II option allows the patentee to buy another 11 months of decision time for a fixed sum which is small compared to the actual filing and translation costs that occur during the actual examination procedure. It is for the same reason that both PCT indicators can be considered as indicators for uncertainty on the side of the patentee which I attribute to a substantial anticipated risk of economic failure of the patent. At the same time, the PCT II indicator may hint at the complete opposite case indicating that the patentee expects significant profits from protecting his invention. This is due to the following reason: The international preliminary examination carried out during the second phase of the PCT (months 19-30 after the filing date) is of real value to the patentee if he/she actually seeks international protection and does not only follow the PCT II to gain more time for his/her decisions. If the applicant is convinced about the economic value of his/her patent from the first day and if he/she wants to protect her invention internationally, he/she will choose the PCT II option and use the preliminary international examination report in the later granting procedure with the EPO or national patent offices. The international preliminary examination report is acknowledged as a substitute for regional or national examination reports by the EPO and several national patent offices, thus the patentee saves time and money by following PCT II in the end if he/she really seeks international protection. Coming back to this study, the empirical findings can now be explained quite well. In this sample of granted patents, PCT II cases should predominantly be cases in which the patentees sought to gain global protection as soon as possible. Thus, the PCT II indicator truly correlates with patent value. The same rationale, however, does not hold true for the PCT I indicator for the reasons given above. In this sample it is presumably instead an indicator of uncertainty at the beginning of the filing process and no significant correlation with patent value can be expected. The results confirm in part the findings by Guellec and van Pottelsberghe de la Potterie (2000), who also find that PCT II should indicate a higher value of a patent application than PCT I. However, the authors give no complete explanation for the phenomenon they observe.

Finally, the differences in the levels of significance for the two acceleration requests can be explained very quickly. While the accelerated search request does not involve any costly decision or commitment by the applicant, the accelerated examination request involves some commitment by the patentee as mentioned above (bivariate analysis). It is therefore not surprising that the last indicator is significantly correlated with the likelihood of an opposition whereas the first one turns out to be insignificant. Column B of Table 6 shows a specification in which the individually and jointly insignificant coefficients of specification 6A were dropped. By showing the marginal effects of specification 6B, column 6C conveys an impression of the orders of magnitude of the different effects. The strongest effects are for the PCT II indicator. PCT II applications in this sample are 87% more likely to be involved in an opposition than other patents. Also the accelerated examination request indicator is very strong. Patents which were examined in an accelerated procedure are 41% more likely to be attacked by opposing parties than other patents. It may be concluded that hypotheses H1 is preliminarily confirmed by the data.

Having discussed the procedural indicators, the next three regressions are based on text indicators. The estimations use a probit model with correction for heteroscedasticity as proposed by Harvey (1976)[12]. I chose the heteroscedastic probit model for two reasons, a theoretical and a statistical one. Theoretically, the interviews with the patent attorneys pointed at the problems of differing *individual modes of drafting* patent applications leading to systematic noise in the data across various applicants. Statistically, models 7A through 7C support this assumption.

Table 7 shows the results of three different regressions of the opposition variable on text indicators.

Insert Table 7 about here

Column 7A models the likelihood of an opposition using all text indicators computed in the study. The upper part of Table 7, column A, shows the first regression results of the most comprehensive specification in which all explanatory variables are used to model the likelihood

[12] Appendix B shows the likelihood function.

of an opposition and the *variance* of the dependent variable at the same time. The lower part of Table 7, column A, shows the respective auxiliary regression. Here the coefficients describe the correlation between the explanatory variables and the variance in the main regression. Following an approach published by Lechner (1991) I carried out joint tests of significance for the individually insignificant coefficients in the auxiliary regression 7A to arrive at a robust specification for the auxiliary regression containing only few variables. The result is presented in the lower part of column 7B. The upper part of column 7C finally presents a specification modeling the likelihood of an opposition with correction for heteroscedasticity using only few variables.

As can be seen from column 7C, four of the text indicators correlate significantly with the likelihood of an opposition when correcting for heteroscedasticity, namely the number of words describing the technical problem, the number of technical preferences, the number of independent product claims, and the number of dependent product claims. For a variety of reasons mentioned above, these results should be interpreted carefully. Still, the findings are very plausible given the state of knowledge developed in this paper. First of all, it was carefully predicted in Table 3 that the likelihood of an opposition should increase with the length of the problem description. Assuming the indicator operationalizes the degree of *inventive activity*, then the length of the problem description will correlate positively with the patent's value. This has a positive effect on the likelihood of the opposition. On the other hand, the opponent's expectations of winning the opposition case should fall with rising *inventive activity.* Thus, the result in 7C suggests the following: the positive effect on the likelihood of an opposition due to increased patent value exceeds the negative effect on the opposition likelihood due to the opponent's diminished expectations of winning the case. Hence, the number of words describing the technical problem mainly correlates with the potential profits from protecting the invention.

Moving two coefficients down in column 7C, there is a positive correlation between the number of independent claims and the likelihood of an opposition. Again, the result seems plausible. Assuming that independent product claims operationalize the *breadth* of the patent as suggested in Table 3, then profits from the patent's value should rise with the number of independent claims. If, as also suggested in Table 3, independent claims are also a measure of the *inventive activity*, then the opponent's expectation of winning the opposition case should fall at the same time. Hence, the result found in 7C suggests that the same rationale applies to independent claims as to the number of words describing the technical problem. The positive effect on the likelihood of an opposition due to increased patent value exceeds the negative effect on the opposition likelihood due to the opponent's diminished expectations of winning the case. Hence, the number of independent claims should correlate with the patent's value.

Theoretically, the last rationale could simply be applied to the *dependent* product claims, too. However, a more elaborated line of argument seems more convincing. As proposed in Table 3, dependent claims often serve as so-called 'fall-back' options. It seems plausible to assume that patentees are more likely to insert those costly fall-back options into a patent application when they face a higher chance of being attacked. Again, they do face a higher chance of being attacked the more valuable the patent and the higher the opponent's anticipated probability of being successful in an opposition. Given the fact that the patent attorney faces additional costs for inserting each additional dependent claim, there is good reason to believe that also dependent claims are a valid correlate of a patent's value.

Finally, I would have expected the likelihood of an opposition to rise with the number of technical preferences. Table 3 suggests that technical preferences either serve as fall-back options or that that they reflect market proximity. However, in both cases the likelihood of an opposition should not drop. Hence, the results found in 7C conflict with the theoretical expectations and additional explanations are needed. In fact, when looking again at the data in detail I find the highest number of technical preferences in those cases where the description of the technical invention is most comprehensive. This leads to the assumption that technical preferences possibly operationalize the *disclosure* of the patent rather than anything else. Assuming that technical preferences are an indicator for the *disclosure* of the patent, the results become then very plausible with respect to Article 83 EPC. According to this article, opponents may substantiate their opposition by blaming the patentee of insufficient *disclosure*. Then, the result in 7C suggests that the patentee invested additional time in the draft of the patent and in carrying out further experiments to reduce the likelihood of a substantiated opposition. His/her willingness to incur extra costs point at a high expected value of the patent. Apparently, however, the negative effect on the likelihood of an opposition mediated by decreased expectations of the opponent of winning in the opposition outweighs the positive value effect on the likelihood of the opposition. Overall, it can be concluded that the data provide some preliminary empirical evidence for the validity of hypotheses H2.

Finally, Table 8 shows two further regressions. In column 8A the likelihood of an opposition is modeled using both procedural indicators and text indicators at the same time. Column 8B shows the corresponding marginal effects.

Insert Table 8 about here

Surprisingly, in this joint model of procedural and text indicators, only the number of application claims turns out to have a significant coefficient among all the text indicators. Again, the explanation of this result may only be preliminary and was not necessarily to be expected according to Table 3. It seems, however, as if patents protecting applications are less valuable than patents that protect products which is consistent with the notions in the interviews. Comparing specifications 6 and 8, I find that the number of application claims adds to the explanatory power of the procedural indicators. This result gives some first empirical evidence for hypotheses H3.

Conclusions and Outlook on Future Research

Recalling the scope of the paper, I finally ask as to what extent the paper enhances our understanding of measuring patent value with indicators. This paper claims to contribute to a better understanding in three different ways. First, it expands the theory by analyzing patent attorneys' work which appears crucial to interpret observable legal actions as indicators of patent value. In particular, new ideas for the compilation of various new indicators arise from the descriptions of patenting strategies. The paper suggests that operationalizations of key variables in the filing process, such as the *state of the art of existing technology*, the *inventive activity*, and the *breadth* of the patent should be suitable value indicators. Secondly, the paper confirms convincingly the validity of a new procedural indicator of patent value, the so-called accelerated examination request. The data provide empirical evidence that patents appear more valuable when patentees are willing to make a *cost commitment* early during the filing procedure. At the same time it reconfirms the validity of other procedural indicators of patent value, offering more detailed explanations for their validity than described in the literature so far. Third, the paper provides some preliminary results on the appropriateness of text indicators as additional measures of patent value. The results suggest that new indicators of patent value can be found by counting the number of words describing the technical problem, the number of technical preferences, independent and dependent product claims, and application claims. The results of this exploratory study furthermore suggest that procedural indicators and text indicators add up in their explanatory power. From an applied standpoint, the two following remarks may of the highest interest: First, all of the tested indicators in this study are available early in the life-time of a patent, draw from publicly available information and are computable at low cost. Secondly, specification number 8 consisting of procedural and text indicators predicts the occurrence of an opposition correctly in 90% of all cases. Thus, in cases where the value of a patent portfolio roughly equals the sum of its individual patents the indicator approach works with considerable precision. This second finding has an important managerial implication. It suggests that indicator methods may become interesting patent valuation tools for analysts and R&D managers also at the corporate level.

Despite the variety of results listed above, however, the paper only claims to provide preliminary empirical results that are worth of further investigation in the future. As laid out in detail in the discussion of the multivariate results, some limitations of the research design used in this study could be avoided in future work. Ambiguities in the interpretation of individual indicators will diminish if the actual patent value is used as the dependent variable. Corrections for heteroscedasticity should become obsolete if the sample consists of patents from only one company. As mentioned, however, surveys of the last sort are very costly and time consuming and require justification by preliminary results as shown in this study.

Finally, I see an alternative chance to extend on this work in the future that does not require costly primary data as suggested in the previous paragraph. Structural models of patent litigation using information on opposition outcomes appear more appropriate to validate indicators of patent value unambiguously than the research design in this study. But, then again, we must learn to walk before we run.

References

Albert MB, Avery D, Narin F, McAllister P. 1991. Direct Validation of Citation Counts as Indicators of Industrially Important Patents. *Research Policy* **20**: 251-259.

Cockburn I, Griliches Z. 1988. Industry Effects and Appropriability Measures in the Stock Market's Valuation of R&D and Patents. *American Economic Review* **78**(2): 419-432.

Cohen WM, Nelson RR, Walsh JP. 2000. *Protecting Their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (or not)*. NBER: Cambridge MA.

Conolly R, Hirsch B, Hirschey M. 1986. Union Rent Seeking, Intangible Capital, and Market Value of the Firm. *Review of Economics and Statistics* **68**(4).

Conolly R, Hirschey, M. 1988. Market Value and Patents: A Bayesian Approach. *Economics Letters* **27**(1): 83-87.

EPO .1998. Annual Report. European Patent Office: Munich.

Griliches Z. 1981. Market Value, R&D, and Patentes. *Economic Letters* **7**: 183-187.

Hall B, Ham, R. 2001. The Determinants of Patenting in the U.S. Semiconductor Industry, 1980-1994, *RAND Journal of Economics* **(32)**: 101-128.

Hall BH, Jaffe A, Trajtenberg M. 2000. *Market Value and Patent Citations: A First Look*. NBER: Cambridge MA.

Harhoff D, Reitzig M. 2000. *Determinants of Opposition Against EPO Patent Grants: The Case of Pharmaceuticals and Biotechnology*. CEPR: London UK.

Harhoff D, Reitzig M. 2001. Strategien zur Gewinnmaximierung bei der Anmeldung von Patenten: Wirtschaftliche und rechtliche Entscheidungsgrößen beim Schutz von Erfindungen. *Zeitschrift für Betriebswirtschaft* **5**: 509-530.

Harvey A. 1976. Estimating Regression Models with Multiplicative Heteroscedasticity. *Econometrica* **44**: 461-465.

Lanjouw JO, Lerner J. 1997. *The Enforcement of Intellectual Property Rights*. NBER: Boston MA.

Lanjouw JO, Schankerman M. 2000. *Characteristics of Patent Litigation: A Window on Competition*. CEPR: London UK.

Lechner M. 1991. Testing Logit Models in Practise. *Journal of Empirical Economics* **16**: 177-198.

Lerner J. 1994. The Importance of Patent Scope: An Empirical Analysis. *RAND Journal of Economics* **25**(2): 319-333.

Megna P, Klock M. 1993. The Impact of Intangible Capital on Tobin's Q in the Semiconductor Industry. *American Economic Review* **83**: 265-269.

Narin F., Hamilton K., Olivastro D. 1997. The Increasing Linkage Between U.S. Technology and Public Science. *Research Policy* **26**(3): 317-330.

Pakes A. 1986. Patents as Options: Some Estimates of the Value of Holding European Patent Stocks. *Econometrica* **54**(4): 755-784.

Putnam J. 1996. The Value of International Patent Rights. Yale University Press, Connecticut NJ.

Rahn G. 1994. Patenstrategien japanischer Unternehmen. *Gewerblicher Rechtsschutz und Urheberrecht (International)* **5**: 377-382.

Reitzig M. 2002. Die Bewertung von Patentrechten - eine theoretische und empirische Analyse aus Unternehmenssicht: Duncker & Humblot, Berlin.

Rivette K, Kline D. 2000. Discovering New Value in Intellectual Property. *Harvard Business Review* **January-February**: 54-66.

Schmoch U, Kirsch N. 1994. Analysis of International Patent Flows. Fraunhofer Institut für Innovationsforschung und Systemtechnik: Karlsruhe Germany.

Tong X, Frame JD. 1992. Measuring National Technological Performance With Patent Claims Data. *Research Policy* **23**: 133-141.

Trajtenberg M. 1990. A Penny for Your Quotes: Patent Citations and the Value of Innovations. *RAND Journal of Economics* **21**(1): 172-187.

Table 1
Established Indicators of Patent Value

Variable	Validity		Availability in Time	Compilation Costs
	Theoretical Foundation[2)]	Empirical Evidence as of Today[2)]	(Months after Filing Date)	(M): Manual Computation Necessary (E): Electronic Computation Possible
Patent Age	++	-	48+	E
Market Value of Corporation	++	++	-	M, Partially E
Backward Citations	+	+/-	18	E
Forward Citations	++	++	Ca. 42+	E
Family Size	++	+	18+ (preliminary)[3)] ca. 42+ (finally)	E
‚Scope‘	+	-	18+ (preliminary)[3)] ca. 42+ (finally)	E
Ownership[1)]	+	++	18+ (preliminary)[3)] ca. 42+ (finally)	E
Number of Claims	++	+/-	18+ (preliminary)[3)] ca. 42+ (finally)	E
Patenting Strategy	++	+/-	18 respectively 19 + (preliminary)[3)] ca. 42+ (finally)	E
Number of Applicants	+	+/-	18+ (preliminary)[3)] ca. 42+ (finally)	E
Number of Trans-Boarder Research Co-operations	+	+/-	18+ (preliminary)[3)] ca. 42+ (finally)	E
Key Inventors	+	+	18+ (preliminary)[3)] ca. 42+ (finally)	E
Legal Disputes (Opposition i n Particular)	++	+/-	ca.42+ (preliminary)[3)] ca. 49+ (finally)	M, Partially E

Legend: [1)]: Differently computed indicators in different studies
[2)]: --: very weak; -: weak; +/-: medium; +: strong; ++: very strong
[3)] : Information available after publication of application. Information can still change during the granting procedure

Table 2
Indicators Computed From Procedural Information

Variable	**Algorithm**	**Expected Effect on Opposition**	**Data Source**
Backward citations to patent literature	Number of patent references to the state of the art that are actively quoted by the patent	+	REFI [©1]
Backward citations to non-patent literature	Number of non-patent references to the state of the art that are actively quoted by the patent	No prediction[5]	REFI [©1]
Family size ‚Scope'	Logarithm[13] of the number of designated states Number of 4 digit IPC classes	+ +	ELPAC [©2] ELPAC [©2]
PCT I application	Dummy variable taking on the value 1 if the patent was filed via PCT and the period of time between filing date and entry into the regional phase is 20 months or less	+	EPASYS [©3]
PCT II application	Dummy variable taking on the value 1 if the patent was filed via PCT and the period of time between filing date and entry into the regional phase exceeds 20 months	+	EPASYS [©3]
Share of A-classifications among backward citations	Share of backward citations among the total number of backward citations which were considered relevant but not innocuous by the preliminary examiners in the Hague	-	REFI [©1.4]
Share of X-classifications among backward citations	Share of backward citations among the total number of backward citations which were considered potentially innocuous by the preliminary examiners in the Hague	+	REFI [©1.4]
Number of inventors	Total number of inventors	(+)[6]	ELPAC [©2]
Number of applicants	Dummy variable taking on the value 1 if more than one applicant is mentioned in the patent	(-)	ELPAC [©2]
Accelerated search request	Dummy variable taking on the value 1 if a request was filed for an accelerated production of the search report	(+)	EPASYS [©3]
Accelerated examination request	Dummy variable taking on the value 1 if a request was filed for an accelerated examination	+	EPASYS [©3]

Legend: 1): REFI: Commercially available citation data bank provided by the EPO Vienna
2): ELPAC: Commercially available EP patent data bank provided by the EPO Vienna.
3): EPASYS: Procedural databank of the EPO. Not commercially available.
4): There are hints that REFI might be incomplete with respect to the citation classifications; at first sight it seems as if data provided by the www-server http://www.epoline.org was more comprehensive. The last server draws from a databank called INPADOC. The reasons for the inconsistencies are unknown for the time being.
5): The effect can theoretically point in both directions.
6): Effects in brackets are subject to greater uncertainty.

13 The operationalization builds on Harhoff and Reitzig (2000).

Table 3
Indicators Computed From Full-Text Information

Variable	Link / Operationalization	Expected Effect on Patent *Value*	Expected Effect on Opponent's Anticipated Probability of Winning	Expected Observable Effect on Likelihood of Opposition
Number of words describing the state of the art	• ***State of the art / Novelty***	+	Unknown 2)	(+)3)
	• ***(Disclosure)***	+	-	
Number of words describing the technical problem	• ***Degree of inventive activity***: Problem of the protected invention as 'counterpart' to the technical solution and therefore to the degree of inventive activity	+	-	(+)
Number of mentioned technical advantages of the invention	• ***Expected demand:*** Technical advantages as a sign of product- and market proximity of the invention	+	Unknown	(+)
	• Technical advantages as hidden fall-back options	Unknown	+	
Number of technical preferences of the invention	• ***Expected demand:*** Technical preferences as a sign of product and market proximity of the invention	+	Unknown	(+)
	• Technical preferences as hidden fall-back options	Unknown	+	
Number of independent claims	• ***Degree of inventive activity***: Independent claims as concise description of the solution and therefore of the degree of inventive activity	+	-	(+)
	• ***Breadth***	+	Unknown	
	• ***Expected demand***: Product and market proximity	+	Unknown	

Continuation of Table 3:

Variable	**Link / Operationalization**	**Expected Effect on Patent *Value***	**Expected Effect on Opponent's Anticipated Probability of Winning**	**Expected Observable Effect on Likelihood of Opposition**
Number of dependent claims	• See above (independent claims)	+	(-)	(+)
	• Fall-back options	Unknown	+	
Number of process claims [1]	• See above (independent claims)	+	(-)	(+)
	• Fall-back options	Unknown	+	
Number of application claims[1]	• See above (independent claims)	+	(-)	(+)
	• Fall-back options	Unknown	+	

Legend: 1): Process and application claims may also be independent claims. To ensure a better distinction between independent product claims and process- and application claims, however, I decided to count them separately in this study.

2): In these cases the existing theory does not allow to predict an effect .

3): Effects in brackets are subject to greater uncertainty.

Table 4
Descriptive Statistics

Variable	Mean	S.D.	Min.	Max.
Opposition (0: no; 1: yes)	0.13		0	1
Family size (Log. of number of designated states)	2.12	0.45	0	2,83
Number of inventors	3.35	1.66	1	8
Number of applicants > 1 (0: 0; 1.>=1)	0.02		0	1
„Scope“	2.24	1.87	1	13
PCT II application (0: no; 1: yes)	0.03		0	1
PCT I application (0: no; 1: yes)	0.01		0	1
Accelerated examination request (0: no; 1: yes)	0.02		0	1
Accelerated search request (0: no; 1: yes)	0.01		0	1
Backward citations to patent literature	3.05	2.03	0	13
Backward citations to non-patent literature	0.81	1.28	0	12
Share of Citations classified as „A“ among the total number of backward citations	0.42	0.38	0	1
Share of Citations classified as „X“ among the total number of backward citations	0.10	0.24	0	1
Number of words describing the state of the art	326.07	257.19	0[1)]	2115
Number of words describing the technical problem	36.20	29.99	0[2)]	295
Number of technical advantages	4.41	5.69	0	49
Number of technical preferences	35.74	37.72	0	304
Number of independent claims	0.64[3)]	0.58	0	7
Number of dependent claims	2.51	3.25	0	21
Number of process claims	4.28	4.15	0	26
Number of application claims	0.77	1.50	0	11

Legend: N=813

[1)]: In these cases it was impossible to unambiguously identify a passage in the text that was solely referring to the state of the art (see text).

[2)]: In these cases it was impossible to unambiguously identify a passage in the text that was solely referring to the technical problem (see text).

[3)]: As independent process- and application claims were counted separately from independent product claims, the mean of ‘independent’ (product) claims may well be below unity.

Table 5
Bivariate Correlations Between Explanatory Variables and Occurrence of Opposition

Variable	**No Opposition**	**Opposition**	**t-Test[14] (P-Value)**
Family size (Log. of number of designated states)	2.08	2.17	-1.60* (0.10)
Number of inventors	3.44	3.06	1.97** (0.05)
Inventors > 1 (0:0; 1:>=1)	0.01	0.02	-1.23 (0.22)
‚Scope'	2.29	2.11	0.86 (0.39)
PCT II application (0: no; 1: yes)	0.00	0.20	-12.47*** (< 0.001)
PCT I application (0: no; 1: yes)	0.01	0.01	0.02 (0.98)
Accelerated examination request (0: no; 1: yes)	0.01	0.10	-5.91*** (< 0.001)
Accelerated search request (0: no; 1: yes)	0.01	0.02	-1.45 (0.15)
Backward citations to patent literature	3.07	3.10	-0.14 (0.89)
Backward citations to non-patent literature	0.79	0.60	1.34 (0.18)
Share of Citations classified as „A" among the total number of backward citations	0.44	0.35	2.02** (0.04)
Share of Citations classified as „X" among the total number of backward citations	0.09	0.11	-0.44 (0.65)
Number of words describing the state of the art	318.60	336.77	-0.62 (0.53)
Number of words describing the technical problem	35.33	40.79	-1.56 (0.12)
Number of technical advantages	4.41	4.91	-0.74 (0.46)
Number of technical preferences	38.10	36.69	0.38 (0.76)
Number of independent claims	0.62	0.54	1.10 (0.27)
Number of dependent claims	2.40	2.71	-0.83 (0.41)
Number of process claims	4.23	4.10	0.27 (0.78)
Number of application claims	0.75	0.66	0.54 (0.58)

Legend: *: Significant at 10% level (two-tailed test)
**: Significant at 5% level (two-tailed test)
***: Significant at 1% level (two-tailed test)

[14] Asymptotic t-tests for dummy variables.

Table 6
Likelihood of Opposition Modeled by Procedural Indicators (Simple Probit)

Independent Variable	**Column A** **Probit Coefficient (S.D.)**	**Column B** **Probit Coefficient (S.D.)**	**Column C** **Coefficient of Marginal Effect in Regression B (S.D.)**
Family size	0.31** (0.13)	0.32*** (0.13)	0.06*** (0.02)
Number of inventors (Coefficient x 10)	-0.53 (0.41)	-	-
Inventors > 1	0.20 (0.42)	-	-
‚Scope' (Coefficient x 10)	-0.03 (0.36)	-	-
PCT II application(0: no; 1: yes)	2.93*** (0.52)	3.09*** (0.51)	0.87*** (0.05)
PCT I application (0: no; 1: yes)	0.08 (0.55)	-	-
Accelerated examination request(0: no; 1: yes)	1.37*** (0.48)	1.29*** (0.41)	0.41*** (0.16)
Accelerated search request (0: no; 1: yes)	-0.26 (0.54)	-	-
Backward citations to patent literature (Coefficient x 10)	0.62* (0.33)	0.71** (0.31)	0.13** (0.05)
Backward citations to non-patent literature	-0.08 (0.06)	-	-
Share of Citations classified as „A" among the total number of backward citations	-0.20 (0.19)	-	-
Share of Citations classified as „X" among the total number of backward citations	0.05 (0.27)	-	-
Dummy for organic fine chemistry	-0.98** (0.45)	-0.55*** (0.15)	-0.09*** (0.02)
Dummy for polymer chemistry	-0.58 (0.44)	-	-
Dummy for biotechnology	- 0.36 (0.50)	-	-
Dummy for pharmaceutical chemistry	-0.46 (0.54)	-	-
Dummy for petrol industry Basic chemicals	-0.33 (0.45)	-	-
Constant	-1.25** (0.53)	-2.07*** (0.28)	-
Wald χ^2 (A:17 / B:5 / C:5)	83.60 (P<0.001)	71.93 (P<0.001)	71.93 (P<0.001)
Pseudo R^2	0.22	0.21	0.21
N	813	813	813

Legend:
*: Significant at 10% level (two -tailed test)
**: Significant at 5% level (two -tailed test)
***: Significant at 1% level (two-tailed test)

Table 7
Likelihood of Opposition Modeled by Text Indicators (Het-Probit)

Independent Variable	Column A Coefficient in the Main Regression (S.D.)	Column B Coefficient in the Main Regression (S.D.)	Column C Coefficient in the Main Regression (S.D.)
Number of words describing the state of the art (Coefficient x 1000)	-0.21 (0.16)	-0.22* (0.12)	-0.26 (0.20)
Number of words describing the technical problem (Coefficient x 100)	-0.08 (0.22)	0.29 (0.19)	0.38* (0.21)
Number of technical advantages (Coefficient x 10)	0.08 (0.11)	0.10 (0.07)	-
Number of technical preferences (Coefficient x 100)	-0.43 (0.37)	-0.88* (0.54)	-1.63** (0.72)
Number of independent claims	0.16*** (0.05)	0.13*** (0.04)	0.14*** (0.04)
Number of dependent claims (Coefficient x 10)	0.61*** (0.17)	0.53*** (0.10)	0.64*** (0.13)
Number of process claims (Coefficient x 100)	-4.10* (2.18)	-2.86** (1.34)	-
Number of application claims (Coefficient x 10)	0.83*** (0.27)	-0.11 (0.32)	-
Dummy for organic fine chemistry	-1.22*** (0.34)	-0.96*** (0.27)	-0.81*** (0.18)
Dummy for polymer chemistry	-0.42* (0.25)	-0.27 (0.22)	-
Dummy for biotechnology	-0.03 (0.27)	0.09 (0.23)	0.31* (0.17)
Dummy for pharmaceutical chemistry	-0.39 (0.33)	-0.31 (0.29)	
Dummy for petrol industry Basic chemicals	-0.21 (0.24)	-0.06 (0.21)	-
Constant	-0.57** (0.27)	-0.61*** (0.22)	-0.91*** (0.14)
Wald χ^2 (A:13 / B:13 / C:7)	43.52 (P<0.001)	104.30 (P<0.001)	73.94 (P<0.001)
N	813	813	813

Continuation of Table 7:

Auxiliary Regression (ln s^2 Dependent Variable)	***Coefficient (S.D.)***	***Coefficient (S.D.)***	***Coefficient (S.D.)***
Number of words describing the state of the art (Coefficient x 1000)	*0.08 (0.26)*	-	-
Number of words describing the technical problem (Coefficient x 100)	*0.28 (0.31)*	-	-
Number of technical advantages (Coefficient x 10)	*-0.02 (0.15)*	-	-
Number of technical preferences (Coefficient x 100)	*0.47* (0.29)*	*0.81*** (0.30)*	*1.03*** (0.31)*
Number of independent claims	*-0.27** (0.14)*	*-0.31*** (0.11)*	*-0.32*** (0.10)*
Number of dependent claims	*-0.10*** (0.03)*	*-0.10*** (0.04)*	*-0.05** (0.02)*
Number of process claims (Coefficient x 10)	*0.28 (0.24)*	-	-
Number of application claims (Coefficient x 10)	*-1.64** (0.84)*	-	-
Log Likelihood Ratio Test For s^2: c^2 (A:8 / B:3 / C:3)	*25.74 (P<0.001)*	*21.58 (P<0.001)*	*21.21 (P<0.001)*

Legend:
*: Significant at 10% level (two-tailed test)
**: Significant at 5% level (two-tailed test)
***: Significant at 1% level (two-tailed test)

Table 8
Likelihood of Opposition Modeled by Procedural and Text Indicators (Simple Probit)

Independent Variable	Column A Probit Coefficient (S.D.)	Column C Marginal Effect (S.D.)
Family size	0.37*** (0.13)	0.07*** (0.02)
PCT II application (no: 0; yes: 1)	3.21*** (0.57)	0.88*** (0.04)
Accelerated examination request (no: 0; yes: 1)	1.23*** (0.41)	0.38*** (0.16)
Backward citations to patent literature (Coefficient x 10)	0.73** (0.31)	0.13** (0.05)
Number of application claims	-0.09* (0.05)	-0.02* (0.01)
Dummy for organic fine chemistry	-0.59*** (0.15)	-0.10*** (0.02)
Constant	-2.11*** (0.28)	-
Likelihood χ^2 (A:6 / B:6)	67.66 (P<0.001)	67.66 (P<0.001)
Pseudo R^2	0.21	0.21
N	813	813

Legend:
- *: Significant at 10% level (two-tailed test)
- **: Significant at 5% level (two-tailed test)
- ***: Significant at 1% level (two-tailed test)

Appendix A: Industry Branches

Industry Branch	(4 digit) IPC Subclass
Organic fine chemistry	C07C,-D, -F, -H, -J, –K
Macromolecular chemistry / Polymers	C08B, -F, -G, -H, -K, –L C09D, –J C13L
Pharmaceuticals / Cosmetics	A61K
Biotechnology	C07G C12M, -N, -P, -Q, -R, –S
Agricultural chemistry / Food chemistry	A01H A21D A23B, -C, -D, -F, -G, -J, -K, –L C12C, -F, -G, -H, –J C13D, -F, -J, –K
Petrol industry / Basic chemicals	C09B, -C, -F, -G, -H, –K C10B, -C, -F, -G, -H, -J, -K, -L, –M C11B, -C, –D

Source: Schmoch and Kirsch (1994)

Appendix B: The Likelihood Function of the Het-Probit

The following equation shows the likelihood function of a probit that corrects for heteroscedasticity.

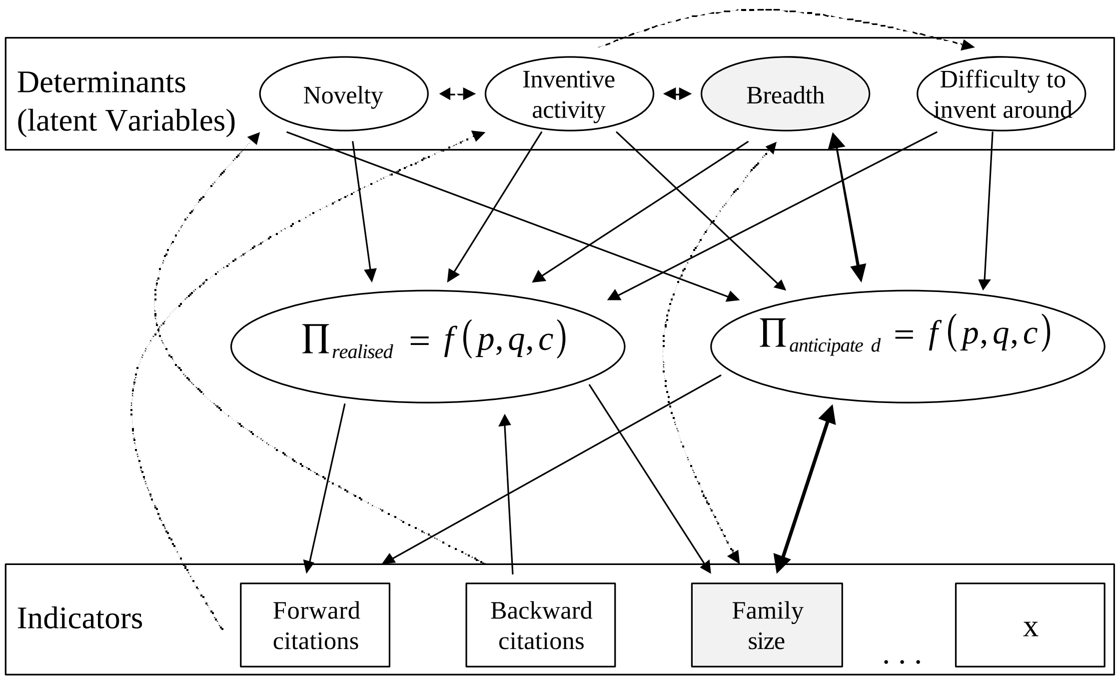
$$\ln L = \sum_{i=1}^{n} \left[y_i \ln \Phi \left(\frac{b\,' x_i}{s_{\,i}} \right) + (1 - y_i) \ln \left(1 - \Phi \left(\frac{b\,' x_i}{s_{\,i}} \right) \right) \right]$$

Here y_i is the observable dependent variable of the main regression and x_i represent the exogenous variables of the main regression. s_i is the standard deviation in the main regression. To correct for heteroscedasticity, the standard deviation (or variance) of the main regression is simultaneously estimated in an auxiliary regression of the following type:

$$s_{\,i} = \exp(g\,' z_i)$$

Here z_i represent the explaining variables of the standard deviation.

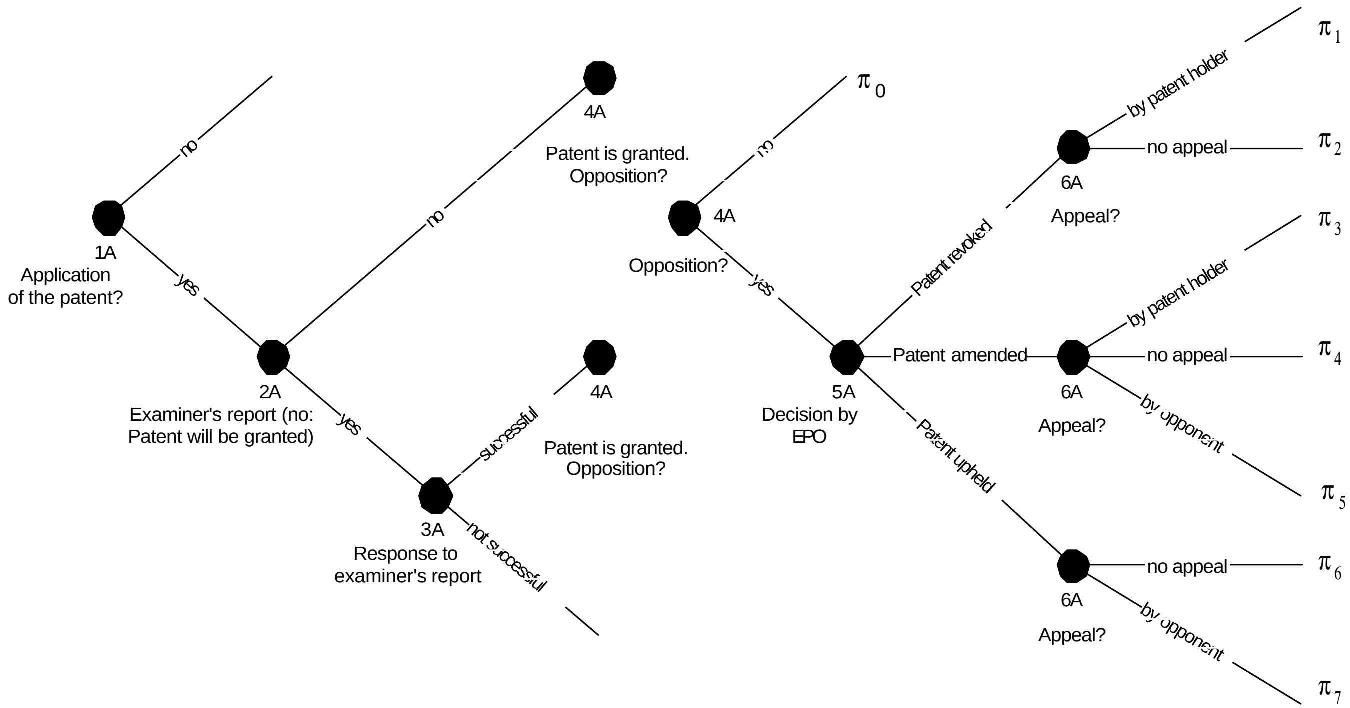
Figure 1: Latent Constructs of Patent Value



Legend: Not all interdependencies are illustrated. The arrows read as follows (examples given below):

- Backward citations are an indicator of the profits from a patent as measured with *ex-post* expertise.
- The number of forward citations is affected by the value of the patent.
- The anticipated value of the patent rises with the family size and *vice versa*
- Grey background: Variable or indicator (partly) *endogenous* for the patentee

Figure 2: The Life Tree of an EP-Patent



Legend: 1A: Patentee's/attorney's decision whether an application is filed or not
2B: Decision by EPO whether to file an examiner's report or not
3A: Patentee's/attorney's response to the examiner's report
4A: Decision by third parties whether to file an opposition or not
5A: Decision by the EPO on the outcome of the opposition plea
6A: Decision by patentee/opponent on appeal

Figure 3: Patent Filing Strategies

Legend: “GebrM”: Official Abbreviation for a German Utility Patent