CERNA WORKING PAPER SERIES

Assessing Indicators of Patent Quality:

Complex vs. Discrete Technologies

Justus Baron
Henry Delcamp

Working Paper 2010:07

Cerna, Centre d'économie industrielle
MINES ParisTech
60, boulevard Saint Michel
75272 Paris Cedex 06 – France
Tél. : 33 (1) 40 51 90 00

May 2010

Assessing Indicators of Patent Quality:

Complex vs. Discrete Technologies

Work in progress

Justus Baron[1], Henry Delcamp[2]

April 30, 2010

Abstract

This article compares indicators of patent quality in complex and discrete technologies using factor analysis and econometric methods. The application of common indicators such as forward citations to complex technologies has repeatedly been put into question. We study the interchangeability of indicators and their capacity to predict litigation on a sample of 9255 patents. Our results do not support the criticism. Even though there are significant differences in the behavior of indicators between samples of complex and discrete patents, issues of complex innovation do not seem to affect the interpretability of quality indicators. Consistently, both forward citations and a compound quality indicator perform equally well for complex and discrete technologies in predicting the likelihood of litigation.

[1] Cerna, Mines ParisTech
[2] Cerna, Mines ParisTech

Introduction

Patents play an increasingly important role in modern economies, and especially in the growing sector of Information and Communication Technologies (ICT). At the same time, especially in the ICT sector, patents are intriguing and puzzling economic objects that have been studied by an extensive literature and give rise to numerous debates. Probably one of the most debated questions is why the number of ICT patents has increased so sharply, and whether this increase reflects a comparable increase in innovation. Many commentators see with suspicion the increasing number of patent files, and concerns about "sleeping patents" of questionable quality and low commercial value are widespread.

Economists traditionally rely upon patent data to measure innovation and technology transfers. This use is challenged by the unequal patent quality and the high number of unused patents. In order to compare the technological quality and commercial value between patents and to single out those patents that are actually used, economic research has come to use various indicators of patent quality. All these indicators are observable characteristics of a patent - like the number of claims or the number of times a patent is cited by posterior patents - that are believed to be driven by an unobservable factor, which is patent quality.

Even though these indicators are now widely used, the validity of this approach is subject to debate and depends upon the concrete area of research. ICT technologies appear to be a particularly problematic field of application for patent indicators. Indeed, economic research has evidenced that patents play a different economic role in ICT industries than in other sectors where patents are important, e.g. in pharmaceuticals or chemistry. Furthermore, firm strategies with respect to patents are different according to the technological field.

These differences have been studied by a large body of economic research. In this literature, ICT technologies are classified as complex technologies, whereas pharmaceuticals or chemistry are classified as discrete technologies. In discrete technologies, single patents protect distinct products that can be brought to the market separately. By contrast, complex technologies are characterized by complementary patents building so-called patent thickets, i.e. dense webs of overlapping patents (Shapiro, 2001). While these complementary patents jointly allow producing a broad range of different products, no single patent enables its owner to produce a distinct product on his own.

Both theoretical arguments and empirical evidence suggest that the distinction between complex and discrete technologies affects the performance of indicators. There are several possibilities how indicator performance could be affected. First, issues specific to complex technologies could bias a specific indicator, i.e. weaken its link to the underlying, unobserved patent quality. In this case, it might be necessary to use different indicators in the assessment of patent quality in complex and in discrete technologies. Second, it is possible that the link between patent quality and all indicators is jointly weakened. For instance, if the various indicators are linked to patent quality, it should be observable that all the indicator variables are driven by a common underlying factor. If no such common factor can be identified, the indicator approach is put into question altogether.

In this article, we assess whether the differences between discrete and complex technologies affect the performance of indicators. We review the empirical literature on indicator performance and on patents in discrete and complex technologies and identify the main concerns against the use of the same indicators in both technological fields. We test these concerns on an extensive database of complex and discrete technology patents using correlation and factor analysis to determine whether the indicators carry a common message and how strongly they converge. Furthermore we use econometric methods to assess how well individual as well as compound indicators perform in explaining the probability that a patent will be litigated. Our analysis is applied to six of the most commonly used indicators of patent quality: forward cites, backward cites, claims, family size, and generality and originality indices.

This analysis reveals important differences between complex and discrete technology patents. Most variable scores are very significantly different from complex to discrete technologies. Overall, the indicators are less correlated among themselves in complex than in discrete technologies. Finally, factor analysis reveals fundamental differences between the samples: whereas in the case of discrete technologies there is only one significant common factor driving the data, the indicators in the case of complex technologies seem to be driven by two rather than by only one factor. Furthermore, we find that the share of the different indicators in the main common factor is significantly different. Nevertheless, both individual as well as compound indicators work well in predicting the likelihood of patent litigation.

These differences could be due to the complementarity of patents, but also to other specificities of the technological sectors. Therefore we introduce a third sample of patents from complex technologies that are disclosed as essential to technological standards.

The use of this sample is an important methodological innovation. It allows disentangling the technological characteristics of patents in complex innovation from the general characteristics of complex technology classes, and especially class-specific firm strategies. As technological standards are meant to achieve interoperability, they are an extreme case of complex innovation. On the other hand, standard essentiality makes sure that there is a commercial application to the patent and thus isolates a sample of patents that are actually used from the mass of sleeping patents.

The second common factor of the indicators is particularly strong for standardized patents, and even outweighs the quality factor. This second factor is largely driven by the generality and originality indices, and opposes backward to forward citations. We thus infer that the fact of finding two rather than one common factor is imputable to the complementarity of patents.

Standards are a particularly rich source of information, as we have detailed data on which patents are complementary and the timing of this interaction (which patents came first, which ones are increments). Thus it becomes possible to directly interpret the second common factor as related to complementarity. For instance, we find that this factor is strongly associated with being a founding patent instead of an increment for a standard.

Finally we assess whether the existence of a second factor affects the performance of indicators in indicating patent quality. Contrariwise to the findings for the random complex patents, the composition of the first factor (the quality factor) is quite similar between discrete and standardized patents. This could indicate that class-specific patent filing strategies are a stronger obstacle to the interpretability of quality indicators than complementarity itself.

The remainder of this article is organized as follows: in the first part, we will review the literature on indicators of patent quality as well as on discrete and complex technologies. This part will discuss the main arguments on how this difference could affect the indicators. In the second part, we will present our data and descriptive statistics. In the third part we will present the results of factor analysis and econometric methods comparing complex to discrete technology patents. The final part of the paper presents the results of the comparison with standardized patents.

I. Theoretical background, literature review

I.1 Indicator performance: benchmarks and survey of the empirical literature

Patent statistics play an increasing role in economic research, as they provide researchers with firm level and sector specific data on innovation output. But the number of patents is not always a good measure of innovation, as patents are heterogeneous. For instance, patents can differ on their commercial value or on their technological significance for further research. Both factors do not only reflect intrinsic characteristics of the patents, but also influences from its technological environment.

For these reasons economists have experimented with various indicators of patent quality to refine the information drawn from patent data. The most commonly used indicators are the number of citations a patent receives by posterior patents (so-called forward citations), the number of claims, and the size of the patent family (i.e. the number of international patent files with the same priority patent) (Griliches, 1990). Other indicators of patent quality include the number of backward cites, i.e. the number of patents cited as prior art and the patent's generality index (measuring the dispersion of citing patents over technology classes). Table 1 summarizes the main indicators of patent quality used in the literature.

Name of the Indicator	Description	Justification
Forward cites	Number of citations received by posterior patents	Indicates the relevance of the patent for further research
Backward cites	Number of citations made to previous patents	Indicates the extent to which the patent makes use of the existing prior art
Number of claims	The number of priority claims made in the patent	Indicates the breadth of the technology claimed by the patent holder
Family size	The number of international patents filed for the same priority patent	Indicates that a patent is important on an international scale, and that its holder is willing to incur high application costs
Generality	Dispersion of cited patents over technology classes	Indicates that the patent draws from various sources, increases the likelihood that the patent is a fundamental rather then incremental innovation
Originality	Dispersion of citing patents over technology classes	Indicates that the patent has been important for a broad field of further research

Table 1 : Patent quality indicators

These indicators are often used indiscriminately in different sectors and to measure heterogeneous phenomena associated with the patents' quality. For example, the number of claims could indicate the breadth of a patent whereas forward cites measure technological significance for further research. Thus, these indicators may be, according to the sector, considered as more or less suited to a study of a specific situation. Consequently, assessing the performance of quality indicators is crucial.

The performance of an indicator could be defined as its ability to explain an economic phenomenon. Therefore, a common way to evaluate indicators of patent quality is to assess how well they predict a specific (costly) decision by economic agents. For example, an indicator that allows predicting, with some accuracy, the probability of a patent to be renewed, licensed, litigated or included into a standard could be defined as a good indicator.

All of these economic phenomena could be affected by factors that are unrelated to patent quality, for instance strategic considerations of the patent holder. These biases need to be borne in mind also for our own evaluation using patent litigation. Nevertheless, as the decisions to license in, to litigate or to renew a patent are costly, they are unlikely to be taken for sleeping patents without any concrete application. As all the different events at least can single out patents that are actually used, it is not surprising that the different evaluations of patent indicators on different events converge.

For instance, the performance of the forward cites indicator has been repeatedly assessed and confirmed. Giummo (2003) finds that patents more often cited are more likely to be licensed, Harhoff et al. (1999) show that patent holders value higher those of their patents that receive more citations, and Hall, Jaffe and Trajtenberg (2005) find that pondering the number of patents by forward cites increases the performance in predicting firm market value. It has furthermore been shown that patents cited more frequently are more likely to be litigated or to be essential to technological standards.

In the light of these reassuring results, economists have carried further their use of patent quality indicators. While these indicators have initially been used as pondering factors, they are increasingly employed as explained variable in inter-patent comparisons. Probably, the most important challenge to the general use of patent quality indicators is the heterogeneity of the patent population. The functions and the mechanisms of patents can vary very much according to external factors, such as the type of patent holder and especially the field of technology.

I.2 Complex vs. discrete technologies

We will structure our analysis around the crucial distinction between complex and discrete technologies. This distinction originates in a paper of Levin & all. from 1987 and has by now been studied by an extensive body of research[3]. A "complex" technology field according to this research is a field where multiple complementary patents, often held by different owners, protect technology necessary for one single product. A typical example is the consumer electronics industry. For instance a BluRay player incorporates several thousand patents held by different major players of the industry. By contrast, in discrete technologies only few patents are directly associated to one product that can be brought to the market independently. Discrete technologies include chemicals or drugs, where one molecular structure or active agent is often protected by few patents, and where in general the IP for one product is held by one single owner.

The functions and the mechanisms of patents can vary very much according to whether they cover complex or discrete technology. There are two dimensions to this divergence. First, the concrete role played by patents depends upon whether there are overlapping property claims or not. A patent that directly confers the right to produce an associated product is a different economic object than one out of many patents in a dense web of complementary patents for the same technology. Second, as demonstrated by previous research (Cohen, Nelson and Walsh, 2000), firm strategies with respect to patents differ from complex to discrete technologies. Most notably, firms patent for different reasons and have a different propensity to patent their inventions.

For several reasons the "complexity" of a technology field could have an impact on the patent indicators. If these indicators are systematically biased by factors regarding their concrete technological field, they are harder to interpret as indicators of quality. For instance the density of the patent web in a complex industry mechanically affects the average number of cites. Independently of its quality, a patent will be cited more often if it covers a technological area where the propensity to patent is high. For the same reason, a patent in such a dense web will have to cite more previous art than a comparable patent in another field of the same technological sector.

[3] Levin et al. (1987), Merges and Nelson (1990), Kusonoki, Nonaka and Nagata (1998), Cohen, Nelson and Walsh (2000), Harhoff and von Graevenitz (2009)

Also firm strategies in the context of complex innovation can bias the indicators. Köhler, Blind and Thumm (2010) find that patents disclosed as essential to technological standards have more claims. Indeed, the existence of overlapping patents could provide incentives to raise the number of claims, as increasing the number of claims increases the chances of the patent to be relevant to future developments of a jointly held technology.

On a different stance, overlapping IP in complex technologies may increase firms' incentives to file numerous patents for few innovations, thereby increasing the size of the families. In complex technologies, many firms use patents for other reasons than excluding their rivals from the use of their technology (Nelson Walsh and Cohen, 2000). Most notably, many firms rely heavily on cross-licensing agreements to cut their way through patent thickets. Hereby patent portfolios play an important role as "mass of negotiation". Furthermore, patent pools and other collective licensing mechanisms reward patent holders according to the number of relevant patents, thereby creating further direct incentives to increase family sizes.

All these factors potentially weaken the link between indicators and patent quality. Important divergences between complex and discrete technologies have been revealed in several empirical analyses of indicator performance. Hall et al. (2005) found forward cites to increase significantly the predictive power of patent counts, but with exception of computer and ITC industries. They argue that when innovations are cumulative, the quality of a patent is in general less correlated with its value. In a different approach, Lanjouw and Schankerman (2004) carry through a factor analysis on four indicators of patent quality. They argue that patent "quality" is the only underlying factor that could jointly affect the number of claims, forward and backward cites and the size of the families. The common factor they identify is mainly driven by forward cites in discrete technologies, and mainly driven by claims in complex technologies.

II. Data and Descriptive statistics

II.1 Construction of the samples and variables

Our first objective was to compare complex with discrete technologies. As discussed, we introduced a third sample of standardized patents in order to disentangle strategic and technological factors of complex technology.

As data are most constrained for standardized patents, we first constituted a database of US patents that are essential to technological standards. This database is derived from patent disclosures at 8 standard setting organizations (SSOs) collected by Rysman and Simcoe and from the websites of seven different patent pools (lists of SSOs and patent pools can be found in the annex 6). It comprises overall 3343 essential patents, out of which 993 are part of a patent pool.

By merging these patent lists with the NBER patent database, we inform the technology classes of 3128 patents and verify that the patents in our database cover technology that is classified as "complex" according to previous literature[4]. The concrete classification of technological sectors into complex or discrete technologies is still subject to debate. In our

[4] See von Graevenitz, Wagner and Harhoff (2009) or Cohen, Nelson and Walsh (2000)

analysis, we will concentrate on clear cut cases of industries that are classified as complex or discrete according to several methodologies. Details on our selection of classes can be found in the annex 4.

41 patents in the database are classified as discrete. Based on the remaining patents, we construct a sample of siblings. These are US patents with the same application year and the same technology class randomly chosen from the NBER patent database. This second sample is what we will call in the following the group of complex, non-standardized patents.

Finally, we build up a third sample of discrete patents. These are patents with the same application years as the patents in the other two samples, randomly chosen from a large range of discrete technology classes in the NBER patent database. The detailed, three-digit technology classes of both the complex and the discrete patent samples can be consulted in the annex.

Overall, we have 9255 patent observations. The NBER patent database yields information on citation flows and other important variables. We inform the number of *forward cites* (including and excluding self-cites), *backward cites* as well as the *generality* and *originality* indice**s**, both building upon citation data. We furthermore retrieve the number of *claims*, the *application year* and the *grant year*. We complete this information on patents using the website of the European Patent Office www.espacenet.com, where we also retrieve the *size of the patent families*. We generate a *grant lag* variable, defined as the difference between grant and application year.

By merging the patent database with our own disclosure database, we obtain the concrete technological standard that 1.509 patents are essential to and the dates of disclosure. If one patent is disclosed as essential to several standards, we retain only the standard of the first disclosure. For every standard, we calculate the mean of the disclosure dates of all essential patents. For every patent, we generate an *age_of_disclosure* variable, defined as the difference between the disclosure date and the mean disclosure date for this particular standard. For the 993 pool patents, we use an earlier database including an *age_of_input* variable, defined as the difference between the date of input of a given patent and the date of input of the first patent in the pool. Even though differently constructed, age_of_disclosure and age_of_input both allow studying the chronological order of patents that are essential for the same technology.

Finally, using the Stanford IP litigation database (www.lexmachina.org), we generate a dummy variable - *litigated* - which gives 1 if the patent has been cited in at least one law suit in the database.

II.2 Descriptive statistics

In this section, we will use the comprehensive database to assess the theoretical arguments on the performance of the patent quality indicators in complex technologies.

In a very first step, we produce descriptive statistics on the scores of the different variables in the different samples. The results in table 2 seem to be consistent with some of the concerns against the general use of indicators: indeed, in line with the hypothesis that the complementary nature of innovation in complex industries drives up citation rates, both backward and forward cite rates are significantly higher in the complex than in the non-complex random sample, whereas the scores for claims are not significantly different, and family size is much bigger in the discrete sample. Furthermore, we confirm previous findings that the litigation rate is indeed higher in complex than in discrete industries.

It is a classical result that the propensity to cite differs across technologies and that the technology class must therefore be controlled for when comparing patents from different industries. This does not necessarily imply that the number of cites can not be used as quality indicator in those classes where the scores systematically diverge. Nevertheless, consistently with the theoretical concerns about the interaction between complexity and indicator performance, we find that among patents in the same (complex) technology field, citing rates are still very significantly higher for those patents that are essential to standards, and thus clearly are part of a cumulative innovation effort. We furthermore reproduce other results from the literature, which already established that these patents have more claims and bigger patent families. All of these differences are statistically significant with the exception of claims.

Interpreting these findings is not straightforward. Are the variable scores higher because complexity biases upwards the indicators, or because essential patents are actually "better" or technologically more significant in the sense intended to be measured by the indicators? It is not apparent from the descriptive statistics whether the important differences in the variable scores are due to a bias or to the well-functioning of the indicators.

		Complete sample			Standardized patents			omplex, non-standardize			Non-complex patents		
		Obs	mean	std	Obs	mean	std	Obs	mean	std	Obs	mean	std
all citations received	allcites	9255	25.403	44.924	3128	44.534	60.965	3000	22.369	37.627	3127	9.1747	16.115
_ excluding self-cites	allnscites	9255	23.089	42.534	3128	39.717	57.800	3000	20.946	36.684	3127	8.5137	15.139
citations made	cmade	5007	9.2892	14.171	1678	11.743	16.355	1587	8.8822	15.396	1742	7.2962	9.6885
number of claims	claims	5907	16.782	15.051	1907	19.502	17.495	1931	15.777	12.929	2069	15.212	14.094
Size of patent family	familysize	8515	15.596	46.377	3133	24.830	63.123	2541	6.5139	17.889	2841	13.537	39.943
generality index	genindex	6677	.34749	.36633	2402	.42989	.34771	2043	.39448	.37355	2232	.21582	.34214
originality index	cgen	5173	.22411	.23610	2063	.24931	.22318	1679	.26272	.24583	1431	.14246	.22255
litigation probability	litigated	9470	.03147	.17459	3343	.06671	.24955	3000	.014	.11751	3127	.01055	.10220
grant lag	grant_lag	9255	2.5983	1.3499	3128	2.5742	1.2193	3000	2.879	1.3865	3127	2.3534	1.3882

Table 2 : Descriptive statistics of indicators

In order to reply to this question, we go one step further in the statistical analysis of our database. The following part will present correlation analysis and factor analysis to compare the samples of complex and discrete technology patents.

III. Are indicators driven by the same underlying factors in complex and discrete technologies?

III.1 The principal factor analysis

Factor analysis is a way to describe variability among observed variables through a smaller number of underlying variables called factors. Factor analysis is similar to the principal component analysis. However, factor analysis is concerned with the common covariance of the variables and estimates how much of the variability is due to common factors. The principal component analysis method allows to realize a variance maximizing rotation of the spaces' variable taking into account the overall variance of the variables and not only the communality.

Thus, the factor analysis uses a large number of observations and reveals common patterns underlying the variables. For example, factor analysis is widely used in medicine to highlight risk factors through medical data observations. Factor analysis is also a method regularly used in political sciences to highlight the unobserved political convictions of surveyed people using their expressed opinion on various societal problems. In economics, factor analysis is used when capturing a common phenomenon is more interesting than analyzing individual variables. For example, it is a method used for a very long time to capture the growth phenomenon of a country[5]. Lanjouw and Schankerman (2004) first used the principal factor analysis to identify an overall patent quality factor through four indicators. A more detailed presentation of this method is available in annex 2.

We first run a factor analysis (with the number of factor constrained to 1) on four indicators frequently used to assess the "quality" of a patent namely the number of forward cites, the number of claims, the number of backward cites and the family size of the patent. This method is similar to the method used by Lanjouw & Schankerman (2004).

Our results on this first factor analysis (presented in annex 3) are very closed to the previous results using the same method. We highlight that the impact of forward cites on the common factor is more important for non complex technologies than for complex technologies. Inversely, the impact of the number of claims is more important in the case of complex technologies. These results are confirmed by the following correlation matrices of indicators for the complex and discrete samples:

Non-complex patents				
	allcites	cmade	claims	family~e
allcites	1.0000			
cmade	0.2591	1.0000		
claims	0.1748	0.1223	1.0000	
familysize	0.2613	0.1826	0.0694	1.0000

Table 3 : Indicators correlation matrix discrete sample

[5] For more information on applying these method to the data on a countrys' growth, see Adelman I. and Taft Morris C., « A Factor Analysis of the Interrelationship Between Social and Political Variables and Per Capita Gross National Product », *The Quarterly Journal of Economics*, Vol. 79, No. 4 (Nov., 1965), pp. 555-578

Complex, non-standardized patents				
	allcites	cmade	claims	family~e
allcites	1.0000			
cmade	0.1114	1.0000		
claims	0.1953	0.2370	1.0000	
familysize	0.0862	0.1092	0.0707	1.0000

Table 4 : Indicators correlation matrix complex sample

Table 3 provides a quite reassuring result: forward citations are quite highly correlated with the other indicators. Also backward cites are correlated with claims and family size. Only claims and family size do not seem to be correlated. The sample of complex patents (table 4) gives a different image: forward cites are much less significantly correlated with backward cites and family size. On the other hand, claims are more strongly correlated with forward and backward cites. Nevertheless, no indicator is correlated at more than 10 % with all the other three indicators. This finding indeed suggests that forward cites as quality indicator carry less information in complex than in discrete industries.

Yet more compelling is the number of factors identified if we do not constrain the number of factors :

Eigenvalue	Discrete sample	Complex sample
Factor1	0.60658	0.49734
Factor2	-0.01655	0.05097
Factor3	-0.13174	-0.10796
Factor4	-0.16029	-0.1981

Table 5 : Number of factors four indicators

As can be seen on the table 5, one factor is sufficient to explain most of the common variance in the case of discrete patents. By contrast, in the case of complex technologies, the number of factors retained is two (as can be seen from the number of factors with positive eigenvalues). This could mean that there is another factor but quality that could impact the indicators in the case of complex technologies. We will analyze into depth this hypothesis with a principal factor analysis based on six quality indicators.

To sum up we can say that our first results are consistent with the literature on the subject in the weight difference of quality indicators between complex and non complex technologies. We can also highlight that for complex technologies, there is more than only one factor that seems to impact the quality indicators. These preliminary results are confirmed by the principal component analysis of non complex, complex standardized and complex non standardized technologies available in annex 1.

We then perform the same principal factor analysis using two new indicators: the originality and the generality of the patent. We also test the variable *grant_lag* defined as the time between the application and the grant date. The *grant_lag* is sometimes used as an indicator of quality or at least as an indicator of technological complexity of the patent (Popp, Juhl and Johnson (2004), Harhoff and Reitzig (2004), Harhoff and Wagner (2006, 2009)). The generality and originality, measured by the number of forward or backward cites between the patent and patents from other technological classes, can get an idea of the patents' interest for broader technological applications (Hall & all., 2001). The following table summarizes the eigenvalues for each factor and each sample.

Eigenvalue	Discrete sample	Complex sample
Factor1	0.65608	0.5183
Factor2	0.28677	0.2174
Factor3	0.09281	0.0426
Factor4	0.00616	-0.03609
Factor5	-0.13273	

Table 6 : Number of factors six indicators

Based on the table 6, we choose a number of factors equal to 2. Indeed, for each sample, a number of factors equal to 2 seem to be the best choice in order to restrict the number of factors (and thus facilitating the interpretation) while keeping the maximum variance. However, as we will see in the next paragraph, the indicators do not impact this second factor in the same way between our two samples. We then perform a principal factor analysis with two factors on each sample. The results of this factor analysis with seven indicators are available in annex 3.

First of all, we can emphasize that the variable *grant_lag* does not seem to have an important common covariance with other variables. So, we will remove this variable from the analysis. This result leads us to believe that the variable *grant_lag* does not appear to be linked to the first common factor. It would seem that unlike other indicators, it does not provide any information on the patent quality (if we accept the assumption that the first factor is a compound factor of overall quality). We will thus proceed to another analysis with a varimax rotation of the factor axes and dropping the variable *grant_lag*. The results are summarized in the following table.

Indicators	Complex patents			Discrete sample		
	Factor1	Factor2	Uniqueness	Factor1	Factor2	Uniqueness
Family size	0.2101	0.0288	0.955	0.3811	-0.0925	0.8462
Claims	0.4192	0.0473	0.822	0.2937	-0.0195	0.9134
Allnscites	0.3026	0.1376	0.8895	0.4443	0.0049	0.8025
Cgen	-0.028	0.3492	0.8773	-0.0816	0.362	0.8623
Genindex	0.1102	0.3303	0.8788	0.0667	0.3605	0.8656
Cmade	0.4037	-0.0142	0.8368	0.3766	0.0593	0.8546
Number of observations		*1172*			*867*	
Chi2		*197.59*			*188.47*	
Prob>chi2		*0*			*0*	

Table 7 : Loadings factor analysis six indicators

Table 7 confirms our precedent result. It seems that there are two main factors underlying these indicators. A first factor is mainly correlated to the number of forward cites, backward cites, claims and to some extent family size. This first factor has already been discussed in the literature (Lanjouw & Schankerman) and named “quality”. We will thus continue to call it this way. We can see that for this first factor, the loadings of indicators are quite different between complex and discrete technologies. Thereby, the number of claims seems to have more impact than the number of forward cites on the first factor for the complex sample. It is exactly the opposite in the case of discrete technologies for which the indicator with the most weight on the first factor is the number of forward cites. This difference has already been noted in the literature (Lanjouw & Schankerman, 2004).

Another important difference is the apparition of a second factor having an important impact on the indicators common covariance. This second factor is mainly linked to the generality and the originality of the patent. For complex patents (as opposed to the discrete sample), this second factor also has significant loadings on the citation indicators. A plausible interpretation would be that this factor discriminates between fundamental and incremental innovations; which could be the reason why it is particularly linked to the generality and originality of the patent but also with the number of cites in the case of complex technologies. We will discuss this hypothesis by performing further tests in the next section.

To conclude we can say that the factor analysis highlights important differences between complex and discrete technologies. Thus, the number of forward cites seems to have less impact for complex technologies than for discrete technologies on the first factor. With our two new indicators (generality and originality), the factor analysis also shows a second factor. The loadings of the indicators on this second factor (especially the number of cites) varies greatly between discrete and complex technologies. In the next subsection, we will analyze if the factors have the same power of prediction (of the probability for a patent to be litigated) for complex than for discrete technologies.

III.2 How well do indicators and factors perform in predicting litigation?

III.2.1 A general explanation of litigation using quality indicators

An appealing way to assess indicators of patent quality is to compare their performance in predicting the likelihood of litigation. Litigation on a patent, like renewals, is an event that helps discriminating between those patents that are actually used and other patents.

In a first step, we will once again compare the performance of the four traditional patent quality indicators forward cites, backward cites, claims and family size. We will successively run probit regressions on our dummy variable *litigated* in the three different samples. In this regression, we control for the age of the patent, the type of assignee and the application year of the patents. In the sample of non-complex technologies, forward citations are the only indicator that significantly predicts the likelihood of litigation. In the sample of complex patents the result is the same for forward cites but backward cites also significantly predicts the likelihood of litigation.

Probit litigated	Non complex patents	Complex patents
Forward cites	0.00934***	0.00524***
	(0.003)	(0.001)
Backward cites	-0.02192	0.00744*
	(0.012)	(0.004)
Number claims	0.01079	0.00325
	(0.006)	(0.009)
Family size	0.00086	0.00321
	(0.001)	(0.002)
Age effect	0.03559	0.02392
	(0.042)	(0.023)
Control Assignee	Y	Y
_Cons	-73.52018	-50.24495
	(84.529)	(45.190)
Number of observations	1521	1381

legend: * p<0.05; ** p<0.01; *** p<0.001
Robust standard errors in parentheses

Table 8 : Litigation prediction traditional indicators

A difference-in-difference estimation (table 9) confirms that the influence of forward citations on the likelihood of litigation is not significantly different from discrete to complex patents.

	Complex vs discrete
Allcites	0.41219
	(0.376)
Claims	4.43981
	(5.357)
Family size	4.68071
	(9.115)
Backward cites	16.68145
	(10.337)

Table 9 : Diff in Diff analysis on litigations
(The complete results are available in the annex 5)

This finding is in line with the literature that already established a positive link between forward cites and the likelihood of litigation. The difference-in-difference analysis further reveals that the link between the quality indicators and the likelihood of litigation does not differ significantly from complex to discrete technologies for any of the main indicators,, which provides us with an argument that traditional indicators of patent quality are viable also in complex industries We will now test the performance of the two factors identified previously to predict litigations in the case of complex and discrete technologies.

III.2.2 Litigations and compound factors

This section aims at comparing the previous results on litigations to the two factors (quality and fundamentality) identified in the second part of this paper. We shall therefore proceed to a probit regression of the variable *litigated* with the compound factors as explained variables. We also introduce the variable *grant_lag* as an explanatory variable because we removed it from the factor analysis. The results are presented in the following table.

Probit Litigated	Non complex sample	Complex sample
Quality factor	0.30377**	0.37110***
	(0.109)	(0.065)
Second factor	- 0.13126	0.56177
	(0.174)	(0.357)
Grant Lag	- 0.21835	0.04027
	(0.157)	(0.143)
Age effect	0.03552	0.04746
	(0.062)	(0.034)
Dummy Appyear control	Y	Y
Dummy Assignee control	Y	Y
_cons	- 2.01619***	- 2.85478***
	(0.521)	(0.545)
Number of obs	*719*	*757*
Wald chi2(8)	*27.36*	*46.14*
Prob > chi2	*0.0006*	*0*

legend: * p<0.05; ** p<0.01; *** p<0.001
Robust standard errors in parentheses

Table 10 : Litigation prediction compound factors

Several insights can be drawn from table 10. First of all, we can highlight that our previous results on quality indicators are confirmed. The probability for a patent to be litigated is higher for good quality patents (based on our overall compound factor). This result is consistent with the existing literature (Lanjouw & Schankerman, 2004, Simcoe, Graham, Feldman, 2009, Lerner, 2009). For the first time, we confirm this result using a compound quality factor and not only individual indicators of patent quality. Individual indicators could be biased by litigations. For example forward cites could increase due to the publicity effect of litigation.

We can also stress that there are no significant differences between the complex and discrete sample. In both cases, the parameter of the compound quality factor is positive and very significant which means that for both samples this factor is linked to the probability for a patent to be litigated.
To conclude this third part, we can underline two main conclusions. The first one is that the structure of the indicators varies between our complex and discrete sample. The second one is that, if we take into account this variation, the performance to predict litigations of traditional quality indicators and of our compound factor is almost the same regardless of the sample analyzed.

In the next section, we will divide our sample of complex technologies in two samples : complex non standardized technologies and complex standardized technologies in order to deeply analyze the previous results and especially the second compound factor.

IV. The introduction of standardized patents to disentangle between sleeping and non sleeping patents

This part is based on three samples. Indeed, we divided the complex sample in two different samples: complex non standardized and complex standardized patents. This division allows us to disentangle between complex sleeping patents and complex patents that are commercially used. In fact, samples of complex technologies are often considered to contain many unused (sleeping) patents. A good way to disentangle between sleeping and non sleeping patents is to use patents that were disclosed in a Standard Setting Organization. Indeed, these patents have a business opportunity and are therefore not sleeping patents. Moreover, the introduction of this sample allow us (by using the date of disclosure) to dissociate between complex fundamental and complex incremental patents and therefore to better interpret our second factor identified previously.

IV.1 The factor analysis

First of all, we run the same factor analysis (with six indicators) than in the precedent section but with two samples of complex technologies. The results are presented in the following table (only for the complex samples) :

Indicators	Standardized complex			Complex non standardized patents		
	Factor1	Factor2	Uniqueness	Factor1	Factor2	Uniqueness
Family size	-0.1235	0.2823	0.9051	0.2101	0.0288	0.955
Claims	0.1107	0.3188	0.8861	0.4192	0.0473	0.822
Allnscites	0.2502	0.3504	0.8146	0.3026	0.1376	0.8895
Cgen	0.4537	0.0078	0.7941	-0.028	0.3492	0.8773
Genindex	0.3854	0.122	0.8366	0.1102	0.3303	0.8788
Cmade	-0.1369	0.2202	0.9328	0.4037	-0.0142	0.8368
Number of observations	*743*			*1172*		
Chi2	*165.42*			*197.59*		
Prob>chi2	*0*			*0*		

Table 11 : Factor loadings with six indicators on two samples of complex patents

The results of this factor analysis are striking. Indeed, if we look carefully we can see that the two factors are inversed compare to the precedent analysis. The loadings between the indicators and the compound factors are also very different between the standardized and non standardized sample. Thus, we can see that the first factor of the standardized complex sample is highly impacted by the generality and originality of the patent and to a lesser extent to the number of forward cites. This factor is actually the second factor identified above and identified as a factor discriminating between fundamental and incremental innovation (we will come back to this interpretation on the next subsection). The second factor (which is the first for the complex non standardized patent) is highly impacted by the four traditional indicators of quality and is the overall quality compound factor already analyzed.
These results are very interesting because it means that there is an underlying factor that impacts significantly the common covariance of the indicators for our standardized complex sample. This factor has a much less important impact on our complex non standardized patent. The following subsection will be dedicated to the interpretation of this factor.

IV.2 Interpretation of the factors

The second (first for our complex standardized sample) common factor jointly affecting all indicators could be a plausible explanation for the divergence in the behavior of indicators across samples and especially of backward cites in the case of standardized patents. The common factor that we interpret as "fundamentality" indeed opposes backward and forward cites, whereas they are jointly driven upward by the factor we suggest to interpret as patent quality. In order to strengthen this argumentation, we will test our interpretation of the factors using variables which are specific to standardization.

As explained in the precedent subsection, a first way to interpret the second factor could be the « fundamentality » of the patent. If this is the case, we expect that the most fundamental and therefore broader and more original patents obtain a higher score on the second factor. We will test this hypothesis by comparing it with different variables from the standardization process allowing us to discriminate between fundamental and incremental innovations.

A first possibility to interpret the second factor would be to compare it to other variables such as the timing of disclosure in the Standard Setting Organization or the timing of input in the patent pool (for pooled patents). In order to do that, we created two new variables, *founding patent pool*, which equals 1 if the patent is a pool founding patent and founding_patent_sso which equals 1 if the patent was disclosed before the average age of patent disclosure in the dedicated Standard Setting Organization. These variables allow us to discriminate between fundamental and incremental innovations. The underlying assumption is that founding patents of a pool or an SSO are more fundamental. We choose to create discrete variables because the best way to capture the opposition could arguably be a discrete rather than a continuous variable. Indeed, a patent is or is not fundamental (or incremental) and there is no scale in the incremental (or fundamental) effect. In order to capture this effect, we thus proceed to the following regressions :

Probit	**Founding patent SSO**	**Founding patent pool**
Fundamentality factor	.24171685*** (0.127)	0.25693* (0.127)
Quality factor	.5337134*** (0.196)	0.50440** (0.196)
Age effect	.08695842* (0.094)	0.16499 (0.094)
Dummy appyear control	Y	Y
Dummy Assignee control	Y	Y
_cons	-173.91463* (187.164)	- 327.86429 (187.164)
Number of obs	*2601*	*369*
Wald chi2(22)	*217.33*	*86.89*
Prob > chi2	*0*	*0*

legend: * p<0.05; ** p<0.01; *** p<0.001
Robust standard erros in parentheses

Table 12 : Founding patent / compound factors

Table 12 shows that both factors are related to being a founding patent. The first factor (quality) has already been defined and interpreted. The result for this factor confirmed that founding patents of a standardization process are of better quality than patents disclosed later in the process (see Baron & Delcamp, 2010). The parameter for the second factor is more interesting. It shows that this factor is significantly linked to being a founding patent of a pool or a standardization project. This could confirm our precedent interpretation that the second factor discriminates between fundamental and incremental innovations.

We will now look at the ability of these factors to predict litigations in the case of complex standardized and non standardized patents.

IV.3 How well do indicators and factors perform in predicting litigation?

IV.3.1 A general explanation of litigations using quality indicators

This parts' aim is to compare the ability of traditional quality indicators to predict litigations on each of our two complex sample (standardized and non standardized). The results are presented in the following table.

Probit litigated	**Complex non standardized patents**	**Complex Standardized patents**
Forward cites	0.00524***	-0.0001438
	(0.001)	(0.898)
Backward cites	0.00744*	-0.00224
	(0.004)	(0.519)
Number claims	0.00325	0.00535
	(0.009)	(0.140)
Family size	0.00321	0.00063
	(0.002)	(0.599)
Age effect	0.02392	-0.04009*
	(0.023)	(0.018)
Control Assignee	Y	Y
_Cons	-50.24495	78.00351
	(45.190)	(36.849)
Number of observations	1381	810

legend: * p<0.05; ** p<0.01; *** p<0.001
Robust standard errors in parentheses

Table 13 : Litigation prediction traditional indicators

As we can see in table 13, the results are very different between standardized and non standardized patents. Forward citations have absolutely no impact on the likelihood of litigation among the (heavily litigated) standardized patents. This result could have two explanations. The first one is that standardization impacts the quality of indicators. The second one is that standardization affects litigation strategies. Comparing these patents to their complex siblings by difference-in-difference estimation reveals that not only forward citations, but also backward citations and family size have a significantly lower influence on the likelihood of litigation among standardized patents:

	Standardized vs siblings among complex
Allcites	- 1.07483***
	(0.287)
Claims	- 6.68749
	(4.582)
Family size	- 0.97805**
	(0.352)
Backward cites	- 1.07071*
	(0.433)

Table 14 : Diff in diff estimation likelihood of litigation

This finding suggests that standardization affects litigation strategies rather than the performance of the indicators. The extreme difference in the litigation rate corroborates this suspicion. In the final section, we will analyze this issue into detail.

IV.3.2 A general explanation of litigations using compound factors

This section aims at comparing the previous results on litigations to the two compound factors (quality and fundamentality) identified in the second part of this paper. We shall therefore proceed to a probit regression of the variable litigated with the principal factors as explained variables. The results are presented in the following table.

Probit Litigated	Complex non standardized	Complex standardized
Quality factor	0.37110***	0.05202
	(0.065)	(0.056)
Fundamentality factor	0.56177	0.24812
	(0.357)	(0.185)
Grant Lag	0.04027	- 0.33736***
	(0.143)	(0.099)
Age effect	0.04746	- 0.01025
	(0.034)	(0.033)
Dummy Appyear control	Y	Y
Dummy Assignee control	Y	Y
_cons	- 2.85478***	19.86666
	(0.545)	(65.061)
Number of obs	*757*	*689*
Wald chi2(8)	*46.14*	*30.4*
Prob > chi2	*0*	*0.0067*

legend: * p<0.05; ** p<0.01; *** p<0.001
Robust standard errors in parentheses

Table 15 : Prediction litigation compound factors

The first result of table 15 is that the quality of the patent does not seem to have an impact on the probability to be litigated for patents in the standardized sample. The only variable having an impact (negative and significant) in this case is the grant lag. As explained in the precedent section, this variable does not seem to be a good indicator of patent quality (the covariance between this variable and other indicators of patent quality such as the number of forward cites or the number of claims is very low or null).

Nevertheless, along with existing research on the grant lag we believe that this variable is a good way of capturing the simplicity of a patent[6]. If we accept this hypothesis, then the above result indicates that what is important in the probability to be litigated is the enforceability of the patent and not, as for other samples, the quality of the patent. This result is confirmed by discussions with professionals specializing in standardization. It can be explained by two main reasons. The first one is that standardized patents are already selected and of better quality than non-standardized patents of the same technological class (Rysman & Simcoe, 2007). Thus, litigation is less indicative of the extent to which a patent is used, as all patents are selected and of good quality.

The other reason is that standardized patents are complex and sophisticated technologies not necessarily understandable by judges. As proof of infringement of one single patent allows obtaining a cease and desist order impeding the use of all complementary patents essential to the same standard, it is rational to file a suit on the patent which is the easiest to explain to the court. This is the first time that this result is highlighted in the economic literature on patent litigations.

To conclude this part, we can say that the introduction of standardized patents confirms the importance of the second factor in case of complementary patents. This complementarity factor is linked to the fact that the patent is a founding patent of a standardization project. The compound quality factor of the standardized patents sample is similar to the compound quality factor of the discrete sample meaning that the quality interpretation is not biased by the presence of complementarities between patents. However, the litigations' behaviours are biased in case of standardization and it is thus difficult to conclude based on our results.

Conclusion

This article is devoted to assessing the performance of quality indicators across technological sectors and especially between complex and discrete technologies. Indeed, there is reluctance to use the same indicators of quality in the case of complex technologies because the particularities of these sectors could distort the indicators' performance. For studying this question, we successively used two different and complementary methods. We use the factor analysis method to assess the interchangeability of indicators and econometrics method to measure the capacity of these indicators to predict the likelihood of litigation between complex and discrete technologies.

While variable scores and correlations lend credit to the arguments against the use of quality indicators in complex technologies, a more detailed factor analysis does not support this skepticism. Odd factor loadings in complex technology classes do not seem to be driven by

[6] Popp, Juhl and Johnson (2004), Harhoff and Reitzig (2004), Harhoff and Wagner (2006, 2009)

complexity, but rather by class-specific patenting strategies. Indeed, standardized patents are more similar to patents from discrete technologies than to random patents from complex technologies. This finding implies that among those complex patents that are actually used, just as for patents in other technology classes, there is a common underlying factor which is mainly driven by forward cites. Correspondingly, we find that forward cites as well as a compound quality indicator have the same explanatory power in predicting the likelihood of litigation in complex and in discrete technologies. The cumulative nature of innovation in complex industries does thus not necessarily bias the interpretation of patent indicators as proxies for quality. Rather, these effects seem to be captured by a second, orthogonal factor, which is largely driven by generality and originality, and which is particularly important for complex patents and even outweighs the quality factor in the case of standardized patents.

Differentiating for the two driving factors of patent indicators may in the future enhance our use of patent statistics. This analysis has shown first tentative research paths in which the interaction between quality and "fundamentality" is used to study patent litigation or standardization. In view of such a use, the implications of complementary innovation for patent indicators need to be further studied. For instance, complementary innovation functions differently in different technological sectors. In order to generalize our results on patents essential to technological standards, it would be interesting to compare patents from different types of processes of complementary innovation. Furthermore, research on patent statistics has singled out further variables that have not been used in our article, but which could improve the dissociation between fundamental and incremental innovations. Of particular interests are data on science-industry interaction, such as statistics on citations of non-patent literature.

REFERENCES

- Baron J. and Delcamp H. (2010), *Strategic inputs into patent pools*, Working paper

- Cohen W., Nelson R. and Walsh J. (2000), *Protecting Their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (or Not)*, NBER Working paper

- Giummo J. (2003), *Should All Patentable Inventions Receive Equal Protection? Identifying the Sources of Heterogeneity in Patent Value*, Discussion Paper, University of California, Berkeley

- Griliches Z. (1990), *Patent statistics as economic indicators : a survey*, Journal of economic literature, Vol. 28, N°4, pp.1661-1707

- Hall Bronwyn H., Jaffe A., Trajtenberg M. (2005), *Market value and patent citations*, RAND Journal of Economics, Vol. 36, pp. 16-38

- Hall, B. H., A. B. Jaffe, and M. Trajtenberg (2001), *The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools*, NBER Working Paper 8498

- Harhoff D., Narin F., Scherer F.M., Vopel K. (1999), *Citation frequency and the value of patented inventions*, The Review of Economics and Statistics, Vol. 81, pp. 511-515

- Harhoff D. and Wagner S. (2009), *The Duration of Patent Examination at the European Patent Office*, Management Science, Volume 55, Issue 12, Pages: 1969-1984

- Kohler F., Blind K. and Thumm N. (2009), *Keep ’Em In, Make ’Em Fit : An Analysis of Patent Filing Behaviour for Essential Patents in Standards*, Working Paper

- Kusunoki K., Nonaka I. and Nagata A. (1998), *Organizational Capabilities in Product Development of Japanese Firms: a Conceptual Framework and Empirical Findings*, Organization Science, Volume 9 , Issue 6

- Lanjouw J. and Schankerman M. (2004), *Patent quality and research productivity: measuring innovation with multiple indicators*, Economic journal, 114 (495). pp. 441-465

- Lanjouw J. and Schankerman M. (2004), *Protecting intellectual property rights: are small firms handicapped?*, The journal of law and economics, 47 (1). pp. 45-74.

- Lerner J. (2008), *The Litigation of Financial Innovations*, NBER Working Paper No. W14324

- Levin R.C., Klevorick A.K., Nelson R.R., Winter S.G., Gilbert R. and Griliches Z. (1987), *Appropriating the Returns from Industrial Research and Development*, Brookings Papers on Economic Activity, No. 3, Special Issue On Microeconomics, pp. 783-831

- Merges R. P. and Nelson R. (1990), *On the Complex Economics of Patent Scope*, Columbia Law Review, Vol. 90, No. 4, pp. 839-916

- Popp D., Juhl T. and Johnson D. (2004), *Time In Purgatory: Examining the Grant Lag for U.S. Patent pplications*, Topics in Economic Analysis & Policy: Vol. 4 : Iss. 1, Article 29

- Rysman M., Simcoe T. (2008), *Patents and the Performance of Voluntary Standard Setting Organizations*, Management Science, Vol. 54, N 11, pp. 1920-1934

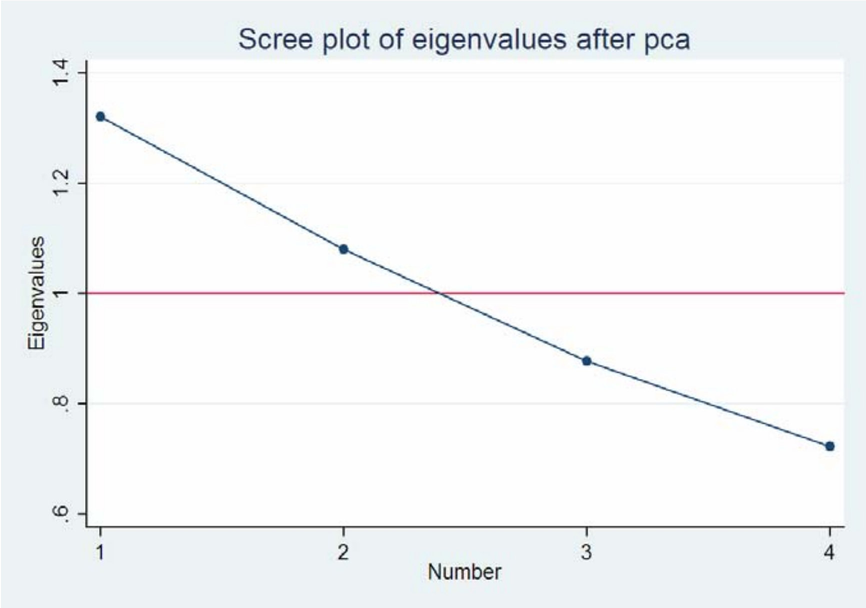
- Simcoe T., Graham S. and Feldman M. (2009), *Competing on Standards? Entrepreneurship, Intellectual Property, and Platform Technologies*, Journal of Economics & Management Strategy, Vol. 18, Issue 3, pp. 775-816

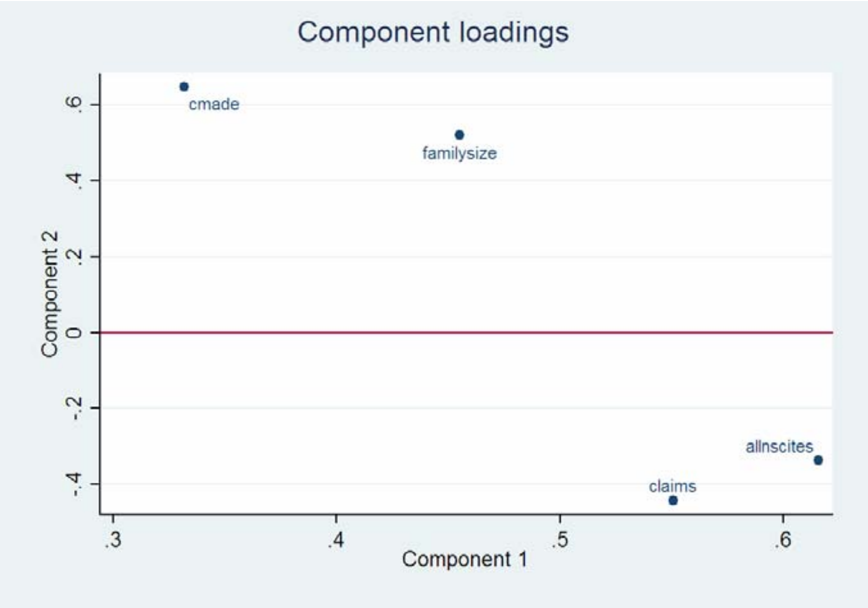
- Von Graevenitz G., Wagner S. and Harhoff D. (2009), *How to Measure Patent Thickets: A Novel Approach*, Discussion paper

Annex 1

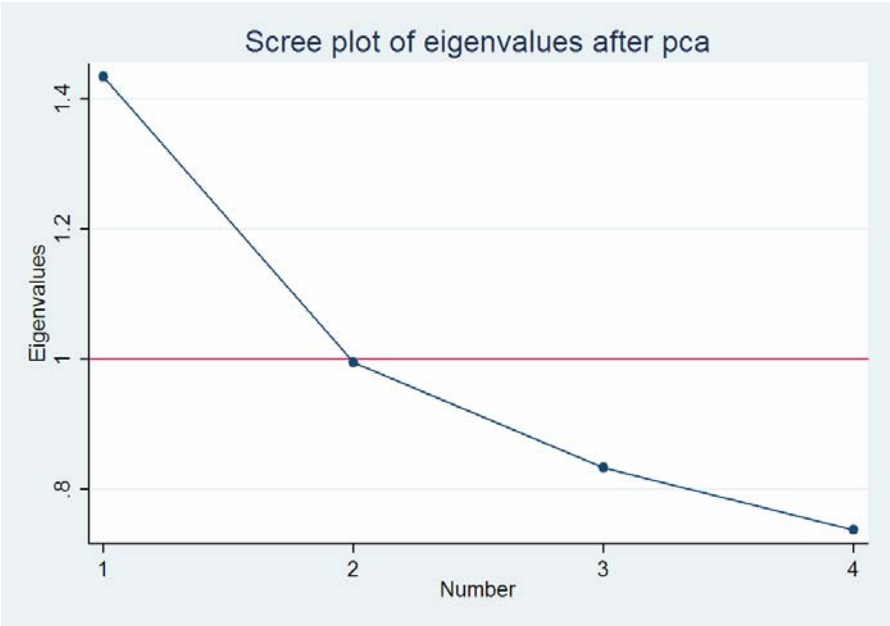
Principal component analysis

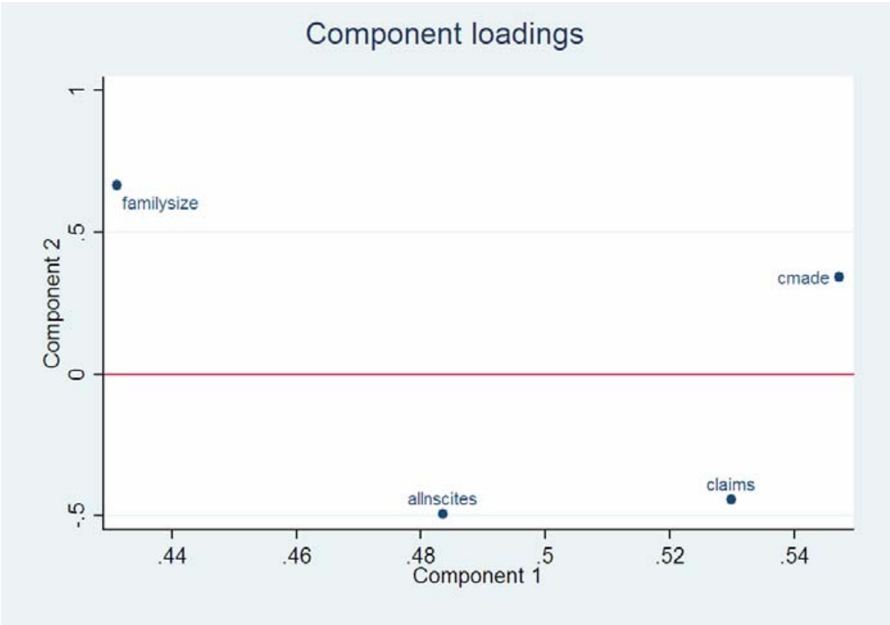
Standardized patents



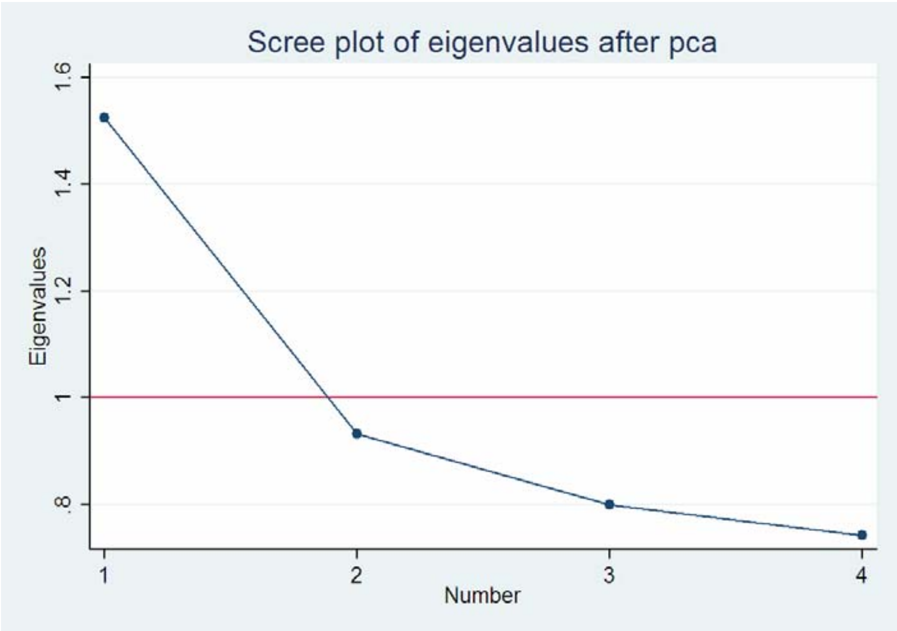


Complex non standardized patents





Non complex sample



Annex 2

The principal factor analysis

For example, if we have a set of n variables $x_1,....,x_n$ that are linearly related to a small number of unobserved factors $F_1,.....,F_n$. Suppose, we have three variables and two underlying factors. Thus, the precedent assumption can be rewritted as follows :

$$x_1 = \beta_{10} + \beta_{11}F_1 + \beta_{12}F_2 + e_1$$
$$x_2 = \beta_{20} + \beta_{21}F_1 + \beta_{22}F_2 + e_2$$
$$x_3 = \beta_{30} + \beta_{31}F_1 + \beta_{32}F_2 + e_3$$

In the factor analysis, the parameters β are called loadings. In order, to conduct a factor analysis, we also need the two following assumptions :

H1 : The error terms e_i are independant of one another and then $E(e_i) = 0$ and $\text{var}(e_i) = \sigma_i^2$
H2 : The unobservable factors (F1 and F2 in this case) are independant of one another and of the error terms and are such that $E(F_j) = 0$ and $\text{var}(F_j) = 1$

Thus, using the above assumptions, we can rewrite the equations of variables explained :

$$x_1 = \beta_{10} + \beta_{11}F_1 + \beta_{12}F_2 + (1)e_1$$
$$x_2 = \beta_{20} + \beta_{21}F_1 + \beta_{22}F_2 + (1)e_2$$
$$x_3 = \beta_{30} + \beta_{31}F_1 + \beta_{32}F_2 + (1)e_3$$

Thus, for the variable x_1 :

$$\text{var}(x_1) = \beta_{11}^2 \text{ var}(F1) + \beta_{12}^2 \text{ var}(F2) + (1)^2 \text{ var}(e_1)$$
$$= \beta_{11}^2 + \beta_{12}^2 + \sigma_1^2$$

Then in the precedent line, we can see that the variance of x_1 can be divided in two parts :

$$\underbrace{\beta_{11}^2 + \beta_{12}^2}_{\substack{Common \\ \text{var}\,iance}} + \underbrace{\sigma_1^2}_{\substack{Specific \\ \text{var}\,iance}}$$

Suppose we take into account two variables and would like to calculate the covariance, then we can write :

$$x_1 = \beta_{10} + \beta_{11}F_1 + \beta_{12}F_2 + (1)e_1 + (0)e_2$$
$$x_2 = \beta_{20} + \beta_{21}F_1 + \beta_{22}F_2 + (0)e_1 + (1)e_2$$

And thus rewrite the covariance of these two variables :

$$Cov(x_1, x_2) = \beta_{11}\beta_{21}\operatorname{var}(F_1) + \beta_{12}\beta_{22}\operatorname{var}(F_2) + (1)(0)\operatorname{var}(e_1) + (0)(1)\operatorname{var}(e_2)$$
$$= \beta_{11}\beta_{21} + \beta_{12}\beta_{22}$$

Annex 3

The factor analysis of four indicators

The following table summarizes the results of a principal factor analysis of the four main indicators of patent quality used by Lanjouw & Schankerman.

Indicators	Non complex sample	Complex non standardized patents	Standardized patents
Family size	0.21431	0.20132	0.19382
Claims	0.19413	0.26642	0.2634
Allnscites	0.2947	0.23164	0.30488
Cmade	0.30157	0.27689	0.13172
Number of observations	*1330*	*1412*	*800*
chi2	*224.15*	*185.25*	*81.09*
Prob>chi2	*0*	*0*	*0*

The factor analysis of seven indicators

Indicators	***Complex sample***			***Discrete sample***		
	Factor1	Factor2	Uniqueness	Factor1	Factor2	Uniqueness
Family size	0.14783	-0.00216	0.9518	0.21734	-0.08186	0.8667
Claims	0.29402	0.01792	0.8292	0.20303	-0.03628	0.889
Allnscites	0.22158	0.09388	0.8763	0.28525	0.00489	0.8077
Cgen	-0.02963	0.29065	0.8777	-0.03578	0.31509	0.845
Genindex	0.0651	0.26953	0.8758	0.0865	0.26361	0.8795
Grant_lag	0.06929	0.02512	0.9871	0.11343	-0.10631	0.9442
Cmade	0.2592	-0.03778	0.8664	0.27025	0.06311	0.8249
Observations		*1199*			*789*	
chi2		*205.62*			*211.67*	

Annex 4

List non complex technology classes

19	Textiles: Fiber Preparation
26	Textiles: Cloth Finishing
28	Textiles: Manufacturing
29	Metal Working
38	Textiles: Ironing or Smoothing
44	Fuel and Related Compositions
57	Textiles: Spinning, Twisting, and Twining
66	Textiles: Knitting
68	Textiles: Fluid Treating Apparatus
71	Chemistry: Fertilizers
75	Specialized Metallurgical Processes, Compositions for Use Therein, Consolidated Metal Powder Compositions, and Loose Metal Particulate Mixtures
76	Metal Tools and Implements, Making
87	Textiles: Braiding, Netting, and Lace Making
99	Foods and Beverages: Apparatus
100	Presses
101	Printing
135	Tent, Canopy, Umbrella, or Cane
139	Textiles: Weaving
148	Metal Treatment
162	Paper Making and Fiber Liberation
164	Metal Founding
228	Metal Fusion Bonding
229	Envelopes, Wrappers, and Paperboard Boxes
423	Chemistry of Inorganic Compounds
424	Drug, Bio-Affecting and Body Treating Compositions
429	Chemistry: Electrical Current Producing Apparatus, Product, and Process
435	Chemistry: Molecular Biology and Microbiology
436	Chemistry: Analytical and Immunological Testing
514	Drug, Bio-Affecting and Body Treating Compositions
518	Chemistry: Fischer-Tropsch Processes; or Purification or Recovery of Products Thereof
585	Chemistry of Hydrocarbon Compounds

List technology classes of standardized patents

Class	Description of the class	Discrete	Complex
8	Bleaching and Dyeing; Treatment of Textiles and Fibers	1	0
16	Miscellaneous Hardware	1	0
29	Metal Working	1	0
36	Boots, Shoes, and Leggings	1	0
40	Card, Picture, or Sign Exhibiting	0	1
73	Measuring and Testing	0	2
75	Specialized Metallurgical Processes	1	0
84	Music	2	0
105	Railway Rolling Stock	1	0
119	Animal Husbandry	1	0
169	Fire Extinguishers	1	0
174	Electricity: Conductors and Insulators	0	3
178	Telegraphy	0	1
188	Brakes	1	0
211	Supports: Racks	1	0
235	Registers	0	14
250	Radiant Energy	0	1
257	Active Solid-State Devices (e.g., Transistors, Solid-State Diodes)	1	0
264	Plastic and Nonmetallic Article Shaping or Treating: Processes	1	0
283	Printed Matter	1	0
315	Electric Lamp and Discharge Devices: Systems	1	0
324	Electricity: Measuring and Testing	0	7
326	Electronic Digital Logic Circuitry	0	4
327	Miscellaneous Active Electrical Nonlinear Devices, Circuits, and Systems	0	1
329	Demodulators	0	1
330	Amplifiers	0	7
331	Oscillators	0	3
332	Modulators	0	1
333	Wave Transmission Lines and Networks	0	2
335	Electricity: Magnetically Operated Switches, Magnets, and Electromagnets	0	1
340	Communications: Electrical	0	73
341	Coded Data Generation or Conversion	0	48
342	Communications: Directive Radio Wave Systems and Devices (e.g., Radar)	0	51
343	Communications: Radio Wave Antennas	0	1
345	Computer Graphics Processing, Operator Interface Processing …	0	13
346	Recorders	0	1
347	Incremental Printing of Symbolic Information	3	0
348	Television	0	102
351	Optics: Eye Examining, Vision Testing and Correcting	0	1
358	Facsimile and Static Presentation Processing	0	99
359	Optics: Systems (Including Communication) and Elements	0	17
360	Dynamic Magnetic Information Storage or Retrieval	0	9
361	Electricity: Electrical Systems and Devices	0	2
362	Illumination	2	0
365	Static Information Storage and Retrieval	0	4
367	Communications, Electrical: Acoustic Wave Systems and Devices	0	1
369	Dynamic Information Storage or Retrieval	0	278
370	Multiplex Communications	0	588
375	Pulse or Digital Communications	0	333
379	Telephonic Communications	0	85
380	Cryptography	0	109
381	Electrical Audio Signal Processing Systems and Devices	0	19
382	Image Analysis	0	87
385	Optical Waveguides	0	4
386	Television Signal Processing for Dynamic Recording or Reproducing	0	225

395	Information Processing System Organization	0	120
401	Coating Implements with Material Supply	1	0
423	Chemistry of Inorganic Compounds	1	0
428	Stock Material or Miscellaneous Articles	6	0
430	Radiation Imagery Chemistry: Process, Composition, or Product Thereof	4	0
434	Education and Demonstration	1	0
435	Chemistry: Molecular Biology and Microbiology	1	0
436	Chemistry: Analytical and Immunological Testing	2	0
438	Semiconductor Device Manufacturing: Process	0	1
439	Electrical Connectors	0	13
455	Telecommunications	0	307
473	Games Using Tangible Projectile	0	1
514	Drug, Bio-Affecting and Body Treating Compositions	2	0
524	Synthetic Resins or Natural Rubbers -- Part of the Class 520 Series	1	0
568	Organic Compounds -- Part of the Class 532-570 Series	1	0
604	Surgery	0	1
606	Surgery	0	1
700	Data Processing: Generic Control Systems or Specific Applications	0	2
701	Data Processing: Vehicles, Navigation, and Relative Location	0	4
702	Data Processing: Measuring, Calibrating, or Testing	0	7
704	Data Processing: Linguistics, Audio Compression/Decompression	0	64
705	Data Processing: Financial, Business Practice, Management	0	1
707	Data Processing: Database and File Management, Data Structures	0	16
708	Electrical Computers: Arithmetic Processing and Calculating	0	4
709	Electrical Computers and Digital Processing Systems: Multiple Computer	0	41
710	Electrical Computers and Digital Data Processing Systems: Input/Output	0	11
711	Electrical Computers and Digital Processing Systems: Memory	0	12
713	Electrical Computers and Digital Processing Systems: Support	0	24
714	Error Detection/Correction and Fault Detection/Recovery	0	75
		41	2904

Annex 5

Diff and diff analysis

Complex vs discrete among non-standardized patents					5729 observations	
allcites	Coef.	Robust Std. Err.	z	P>z	[95% Conf.	Interval]
litigated_~x	.4121865	.3755238	1.10	0.272	-.3238266	1.1482
litigated	.7733005	.2554875	3.03	0.002	.2725542	1.274047
complex	.1112962	.0783064	1.42	0.155	-.0421815	.2647739
gyear	-.0215297	.004019	-5.36	0.000	-.0294069	-.0136525
assignee	1.81e-07	1.04e-07	1.73	0.084	-2.40e-08	3.85e-07
class_cites	.0300635	.0026692	11.26	0.000	.024832	.0352949
_cons	44.88661	8.012398	5.60	0.000	29.1826	60.59062

Standardized vs siblings among complex patents					5722 observations	
allcites	Coef.	Robust Std. Err.	z	P>z	[95% Conf.	Interval]
litigated_~d	-1.074828	.2869105	-3.75	0.000	-1.637162	-.5124934
litigated	1.209788	.2741875	4.41	0.000	.6723904	1.747186
standardized	.6799996	.0376878	18.04	0.000	.6061329	.7538663
gyear	-.043323	.0036483	-11.87	0.000	-.0504736	-.0361723
assignee	2.22e-07	8.43e-08	2.63	0.009	5.64e-08	3.87e-07
class_cites	.027993	.0015466	18.10	0.000	.0249616	.0310244
_cons	88.57468	7.276405	12.17	0.000	74.31319	102.8362

Standardized vs siblings among complex patents					5412 observations	
familysize	Coef.	Robust Std. Err.	z	P>z	[95% Conf.	Interval]
litigated_~d	-.9780464	.3520139	-2.78	0.005	-1.667981	-.2881119
litigated	1.283105	.3099156	4.14	0.000	.6756812	1.890528
standardized	1.349428	.0770302	17.52	0.000	1.198451	1.500404
gyear	.0554772	.0080888	6.86	0.000	.0396235	.071331
assignee	-1.47e-07	1.71e-07	-0.86	0.390	-4.82e-07	1.88e-07
class_fs	.0196475	.0020848	9.42	0.000	.0155615	.0237336
_cons	-109.3425	16.14474	-6.77	0.000	-140.9857	-77.69944

Standardized vs siblings among complex patents					3265 observations	
cmade	Coef.	Robust Std. Err.	z	P>z	[95% Conf.	Interval]
litigated_~d	-1.07071	.4333407	-2.47	0.013	-1.920042	-.2213774
litigated	1.044359	.4201508	2.49	0.013	.2208782	1.867839
standardized	.2692302	.0540057	4.99	0.000	.163381	.3750795
gyear	.0351448	.0065111	5.40	0.000	.0223832	.0479064
assignee	5.87e-07	1.26e-07	4.68	0.000	3.41e-07	8.33e-07
class_cmade	.0565481	.0073978	7.64	0.000	.0420487	.0710476
_cons	-68.81575	12.94931	-5.31	0.000	-94.19594	-43.43557

```
Linear regression                                      Number of obs =     4000
                                                       F(  5,  3994) =    46.91
                                                       Prob > F      =   0.0000
                                                       R-squared     =   0.0546
                                                       Root MSE      =   13.179
```

claims	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
litigated_~x	4.439811	5.357131	0.83	0.407	-6.063155	14.94278
litigated	4.145341	3.341689	1.24	0.215	-2.406234	10.69692
complex	-.8567766	.4574249	-1.87	0.061	-1.753585	.0400315
class_claims	.6223871	.0709997	8.77	0.000	.4831881	.7615862
gyear	.4730284	.0436542	10.84	0.000	.3874419	.5586149
_cons	-938.6608	86.98254	-10.79	0.000	-1109.195	-768.1265

```
Linear regression                                      Number of obs =     5392
                                                       F(  5,  5386) =    20.20
                                                       Prob > F      =   0.0000
                                                       R-squared     =   0.0239
                                                       Root MSE      =   31.316
```

familysize	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
litigated_~x	4.680711	9.115128	0.51	0.608	-13.18863	22.55005
litigated	12.27301	5.721348	2.15	0.032	1.056856	23.48917
complex	-8.386181	.945858	-8.87	0.000	-10.24044	-6.531916
class_fs	.2602817	.0471277	5.52	0.000	.1678923	.352671
gyear	.2912809	.1038699	2.80	0.005	.0876538	.4949079
_cons	-571.9294	207.193	-2.76	0.006	-978.1116	-165.7473

```
Linear regression                                      Number of obs =     3329
                                                       F(  5,  3323) =    11.35
                                                       Prob > F      =   0.0000
                                                       R-squared     =   0.0349
                                                       Root MSE      =    12.54
```

cmade	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
litigated_~x	16.68145	10.33662	1.61	0.107	-3.585329	36.94824
litigated	-1.039264	1.194054	-0.87	0.384	-3.38042	1.301892
complex	-.1266001	.5085939	-0.25	0.803	-1.123789	.8705889
class_cmade	.5182637	.0883539	5.87	0.000	.3450301	.6914972
gyear	.2430547	.0520624	4.67	0.000	.1409772	.3451323
_cons	-481.3625	103.8547	-4.63	0.000	-684.9882	-277.7368

Linear regression

Number of obs	=	3828
F(5, 3822)	=	46.09
Prob > F	=	0.0000
R-squared	=	0.0655
Root MSE	=	14.986

claims	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
litigated_~d	-6.687495	4.582267	-1.46	0.145	-15.67142	2.296428
litigated	8.759695	4.186043	2.09	0.036	.5526029	16.96679
standardized	2.140788	.5062138	4.23	0.000	1.148313	3.133263
class_claims	.8717415	.078802	11.06	0.000	.7172436	1.026239
gyear	.2819561	.0493264	5.72	0.000	.1852475	.3786647
_cons	-562.3067	98.15463	-5.73	0.000	-754.7472	-369.8662

Annex 6

List of pools

- 1394
- DVD 6C
- MPEG 2
- MPEG 4 Systems
- MPEG 4 Visual
- AVC
- DVB-T

List of Standard Setting Organizations

- American National Standard Institute
- Alliance for Telecommunications Industry Standards
- European Telecommunications Standards Institute
- Institute for Electrical and Electronic Engineering
- Internet Engineering Task Force,
- International Organization for Standards International Electrotechnical Commission
- International Telecommunications Union
- Telecommunications Industry Association