Paper to be presented at
the DRUID16 20th Anniversary Conference
Copenhagen, June 13-15, 2016

A fresh look at patent citations

Jurrien Bakker
KU LEUVEN
Department of Managerial Economics, Strategy and Innovation
Jurrien.Bakker@kuleuven.be

Bart Van Looy
K.U. Leuven
Managerial Economics, Strategy and Innovation / INCENTIM-ECO
bart.vanlooy@kuleuven.be

Abstract

This paper uses a large sample of citation and renewal statistics of USPTO and EPO patents to provide additional insights in the relation between patent citations and patent value. A novel approach is used by which the relative contributed value of a patent citation is calculated using an adapted logistic regression. The results of these analyses indicate that this relation depends on the exact document that is referenced, the timing of the citation and the originator citation. Moreover, DOCDB family based citations emerged as a possible optimal citation indicator. Finally, large interaction effects between citation timing and other properties of patent citations were found which shed an additional light on the theorized relations between different patent citations and patent value.

Jelcodes:O32,O34

A fresh look at patent citations

Abstract

This paper uses a large sample of citation and renewal statistics of USPTO and EPO patents to provide additional insights in the relation between patent citations and patent value. A novel approach is used by which the relative contributed value of a patent citation is calculated using an adapted logistic regression. The results of these analyses indicate that this relation depends on the exact document that is referenced, the timing of the citation and the originator citation. Moreover, DOCDB family based citations emerged as a possible optimal citation indicator. Finally, large interaction effects between citation timing and other properties of patent citations were found which shed an additional light on the theorized relations between different patent citations and patent value.

Keywords: patent citations, patent renewal, patent value, patent family

1 Introduction

Measuring innovation is one of the main challenges to the modern empirical economist. Moreover, it is even harder to observe the quality and relative impact of a particular innovation. This is important since the impact from innovations can differ greatly (Scherer and Harhoff, 2000): The impact of innovation can differ from the smallest incremental improvements of a single product to the groundbreaking radical breakthroughs which shape our societies (Dosi, 1982). In order to accurately study innovation, it is important to distinguish these innovations from each other. Whenever innovations are observed by way of patented inventions, several attributes of the patent can be taken into account (Squicciarini et al, 2013). The most used attribute of the patent is the number of times the patent has been referenced by later patents. This is done by computing patent citation indicators which count the number of times a patent, or group of patents, has been cited.

Currently, patent citations are considered to be one of the best validated indicators of patent value, with various validation efforts undertaken by a large body of researchers. These validation efforts of the patent citation indicator have been done in various ways, using different dependent variables and by using different sources of patent citation data. Table 1 provides an overview of the most important validation efforts, their methods and the results that were found.

Insert Table 1 about here

Table 1 shows that patent citations are a significant indicator of patent value: In almost all studies we reviewed, significant relations between patent citations and measures of value were found. This is a substantial achievement considering that so many different dependent variables as well as many different estimation methods were used. However, most results were obtained with few other variables considered and then on small datasets and still find that patent citations explain little variance in their studies. We are not the first authors to note this(e.g. Gay and Le Bas (2005), Gittelman(2008)). Later and larger validation efforts such as Gambardella et al.(2008) and Hall et al. (2005) provide a more nuanced picture: patent citations are still significant but only explain a small part of the variance in the quality indicator, ranging between 2-4% percent. This is perhaps an indication that patent citations are better at differentiating between patents with very similar technical characteristics and are thus less effective at explaining value in larger datasets. Moreover, most validation studies only pick up significant effects for groups of highly cited patents, while patent citations do not appear to perform as well for patents with lower valuation. Patent citations seem an indicator that performs consistent but poorly.

In light of the current (dismal) state of the art of patent citations, it is important reexamine how patent citations actually reveal the value of patent documents. The main problem with patent citations is that there is no direct link between patent citations and patent value, rather patent citations are seen as a useful proxy of several kinds of patent value: In his seminal work Grilliches(1998) stated that patent citations may be able to identify social value by referring mainly to ‘important’ patents in the field, an idea supported by the empirical work of Carpenter et al. (1981). Narin et al. (1989) identifies self-citations, which are citations from the same applicant, as indications that the patent is part of an ongoing project. The seminal work of Jaffe et al (1992) contributed to the literature by linking patent citations to knowledge flows. Finally, large scale studies like Gambardella et al. (2008) and Hall et al. (2005) also confirmed that patent citations explain private value.

The interaction between the type of citations and the explanatory power of the indicator is less well understood. This is however, an important issue since strong divergences have been found between citations that are delivered by applicants and those that come from patent examiners (e.g. Alcacer and Gittelman (2006); Hegde and Sampat,2009). The difference between self-citations and citations that come from other patents has also been studied (e.g. Bessen (2008), Thomas(1999)). However, these studies only give a limited insight in the mechanics that govern the explanatory power of patent citations.

With the advent of structured databases with more informative content such as the PATSTAT database, other questions have also emerged. First, it became easier to correct for similar filings in other jurisdictions, using patent families as proposed amongst others by Webb et al. (2005) and used by, amongst others, Gambardella(2008). This brings with it the question of how this addition changes the make-up of the citation indicator, with Bakker et al. (2016) indicating a substantial differences between family corrected and non-family corrected indicators. Second, EPO patents from the PATSTAT databases provide more information on their backwards citations. This opens up the interesting avenue of examining whether the reason of citing is of influence on the citation indicator.

When citation indicators were introduced, their main focus was on the perceived addition to explain social value. Currently most empirical studies use citations to represent innovative quality, or alternatively, the private value that is represented in the invention. Therefore it is important to re-evaluate the relation between citations and private value. In this paper , we will observe the relationship between various patent citation measures and patent value as observed by patent renewal. Our observations will help shed light on how patent citations reveal value in patent documents, as well as provide useful information for researchers that use patent citations as a measurement of (private) value.

2 Methods and Data

2.1 Using patent renewal as a measure of private value

In this paper we will use patent renewal as an indicator of patent value. Patent renewal as an indicator is derived from the process in which the owners of a patent need to pay periodical fees to keep a patent in force. For patents at European offices this fee is paid yearly(after an initial free period of a few years), while owners of USPTO patents only have to pay this fee at 4,8 and 12 years after application. The value of patent documents can then be assessed by observing if any renewal fees have been paid and for what period they were paid. It is to be noted that patent renewal is limited by the maximum lifetime of patents of 20 years. The patent renewal indicator is often seen as an indicator of value (e.g. Pakes and Schankerman, 1984; Pakes, 1984; Lanjouw et al. 1998,Harhoff et al. 1999, Thomas 1999, Hegde and Sampat, 2009) as it reflects an economic decision of the owner of the patent. Thus patent renewal can be thought of as revealing the private value that the owner attributes to the patent.

We choose renewal as an indicator of patent value because it is an indicator that covers virtually all patent applications. Moreover, it is also the only disaggregated (i.e. on the basis of a single patent) that reflects revealed value. All other dependent variables that are used to proxy value[1] and that use a fine grained measure to validate patent citations suffer from two major drawbacks: First, they are costly or impossible to obtain for large sample sizes; second their measure cannot be directly thought of as *revealed* value. They either represent a social construct (e.g. radicalness, creativity) or they use an indirect way of observing private value such as a survey design. Patent renewal is the only indicator of patent value that does not suffer from these two objections.

2.2 Methods for estimating the effect of patent citations on renewal

Patent renewal can be modelled in several ways. Pakes(1986) highlighted the a real option approach by considering that renewing a patent does not only extend patent protection for a limited time but also provides the option of future extensions. This approach has been modelled by (Maurseth, 2005) using survival analyses. Patent renewal can also be modelled using binary approaches by considering for each patent whether it had been renewed up until that benchmark or not. Each approach has its advantages: Survival analyses and capture more variation. However, they have to rely on more assumptions on the underlying model. The main assumption being that patents with different citation rates have similar hazard functions. In later analyses we show that this assumption is often violated. Binary analyses create a simpler model that uses less of the available variation in the data but provides better metrics to understand model fit.

In this paper we decided to use binary analyses. This choice is primarily motivated by the necessity of depicting an understandable model fit in some analyses of this paper. Furthermore, we believe that given the vast amount of patent data available, the added information of different survival times from survival analyses does not substantially benefit our analyses.

[1] i.e. the variables used in Albert et al. (1991), Arts et al.(2013), Carpenter et al.(1981), Gambardella et al. (2008) Harhoff et al. (1999)

In this paper we employ logistic(logit) regressions to estimate whether a patent has been renewed up until its maximum lifetime. By doing these analyses, the explanatory value of patent citations with respect to private patent value can be evaluated. This is a common practice in some of the literature (e.g. Hegde and Sampat 2009). However, in this paper not only the common citation indicators are evaluated. patent citations will be evaluated by their origin and source and more importantly compared to other patent citations to the same documents.

There is a major issue with comparing indicators using logistic regressions and that is that the coefficients of the regression are a biased representation of the 'true' coefficient: the coefficients are biased with respect to the unexplained variance in the model(Allison, 1999). We therefore opt for a model that revolves around relative weights of citations. Next to eliminating the issues with comparing several logistic regressions, it is also an approach that is not dependent on the variation of the distributions of the independent citation sources that we wish to compare.

In order to accurately compare the relative weights of citations with one another, they need to be present in the same regression. When a linear relation between citations and value is assumed, this can be simply done by entering specific counts of citations separately in the regression and estimating the coefficients. For a logistic regression this would take the following form:

$$P(renewal = 1) = P(\varepsilon > C - \vec{\omega}\overrightarrow{Cit} - \vec{\beta}_{control}\vec{x}_{control})$$

Here the left side of the equation represents the chance a patent is renewed, ε a random error with an assumed logistic distribution, *C* a constant, $\vec{\omega}\overrightarrow{Cit}$ denotes the citation vector with different weights and $\vec{\beta}_{control}\vec{x}_{control}$ gives a set of controls with their respective coefficients. Thus by using this equation it is possible to determine the relative weights, and therefore relative importance of citations from each citation source.

Unfortunately citations do not scale linearly with value. Recent papers (Bakker and Van Looy,2015) indicated that the relation between citations and patent value can be best approximated using a loglinear relation. We tested this in our analyses and found that a loglinear relation indeed gives better results. Therefore we estimate the following equation using an adapted logit regression.

$$P(renewal = 1) = P(\varepsilon > C - \beta_{cit}\ln(1 + \vec{\omega}\overrightarrow{Cit}) - \vec{\beta}_{control}\vec{x}_{control})$$

Unfortunately it was not possible to estimate the citation coefficient β_{cit} together with the weight vector $\vec{\omega}$. Therefore we first estimated β_{cit} by setting all weights equal to 1 or $\vec{\omega} = I$. We then used the estimate of β_{cit} to estimate $\vec{\omega}$. As a robustness check we also estimated $\vec{\omega}$ by setting $\beta_{cit} = 1$. This did change some of the relative weights, but not enough to alter the conclusions we reach to in this paper.

2.3 Data

We use patent citation indicators constructed from the October 2013 version of the EPO PATSTAT database. We used this data source not only for information on individual patents but also to extract data on patent families, most notably the DOCDB and INPADOC definitions[2]. We obtained renewal data from the INPADOC_PRS file provided by the 2014 April version of the EPO PATSTAT database. We observed patent renewal for all granted patents of both EPO and USPTO by examining renewal fees paid to the respective office[3]. In the case of EPO patents, we considered a patent renewed if it has been renewed at least one national office, which subscribes to the EPO regional system. This method follows the Single Renewal Approach (SRA) as discussed in van Zeebroeck(2011). In appendix A we will provide more insight in renewal at the EPO as a measure

There are a number of data issues that need to be taken into consideration: First, renewal data suffers from severe censoring as we cannot observe potential renewal after the maximum time (usually 20 years) of the patent. Since we opted for binary analyses where we examine if the patent has been renewed up until its maximum lifetime, this censoring is of less a concern, but it comes into play for the robustness analyses of appendix A. We also observe censoring at the end of period of observation with patents that are still in force at this moment. Patents that suffer from this censoring have been removed from the analysis.

We want to be able to compare renewal analysis at both the EPO and the USPTO. Therefore the sample of patents was constructed using DOCDB families. All members of this family have the same technical content (Albrecht et al. 2011). Our sample contains all DOCDB patent families that have at least 1 USPTO member and 1 EPO member. It is to be noted that this decision biases the dataset to include more valuable patents, as patents with larger families have been found to be of higher value (e.g. Harhoff et al. 2003). We further restricted the families to have granted patents in both EPO and USPTO, since patents cannot be renewed if they were never granted. Finally, we applied basic data cleaning to avoid problems with the USPTO patents due to a rule change in 2001 and the problems of artificial applications in the PATSTAT database.

The sample was further restricted to include only patents that were applied later than 1980 for EPO applications and 1981 for USPTO applications because we are unsure of the accuracy of renewal data of earlier patents[4]. Since renewal for a EPO office is a yearly decision, only patents that are at least 20 years old were used in the analyses (i.e. having application years up until 1993). USPTO renewal only occurs at three stages of the patent life time with full renewal decided at 12 years after application. Therefore USPTO patents up until 2001 are included, giving the USPTO analyses more observations than their EPO counterparts.

We performed the analyses using a basic set of control variables. These variables are both to rule out spurious correlation between general information of the patent, patent renewal and patent citations, and also reflect any easy to obtain knowledge about the patent portfolio of any patent holder. Therefore

[2] The DOCDB family is perceived as a family of technologically equivalent patent documents in different offices, whereas the INPADOC family is (often) larger and is thought to represent the underlying invention better. For exact definitions of these patent families see Albrecht et al. (2010) and Bakker et al. (2016).

[3] i.e. we observed the last registered renewal payment of the patent at its respective office.

[4] We observed a larger number than expected of never renewed patents for earlier dates. This indicates that renewal data may be incomplete, thus making an analysis of these years unreliable.

this analysis can also be used to assess the added value of including the patent citation indicator. We included controls for both the year and the technological area (IPC3)[5] in which the patent was filed. Whenever multiple classes were encountered on the patent, partial counts were used. This was done to make the controls reflect the actual technology class as much as possible, given the level of aggregation that was used.

Furthermore, we also included controls reflecting attributes of the applicant of the patent[6]. We included these dummies because we assume that different applicants likely write different kinds of patent documents, thus affecting their citation rates and also have different evaluation criteria of the renewal decision of a patent. Moreover, smaller applicants might qualify to pay lower maintenance fees at the USPTO (2016). As controls we included the following: the type of applicant, the size of the applicant, the experience of the applicant and the residence country of the applicant. We also included a dummy indicating whether there were multiple applicants. Whenever there were multiple applicants we observe this in a dummy variable and adjust the applicant related variables in the following way. For continuous variables, we defaulted to the oldest and largest applicant as we assume that it is the most experienced actor that will decide on patent renewal. As for the categorical attributes(i.e. applicant type and country of residence), we used a partial count in these cases. The exact definitions of the controls are listed in Table 2. Table 3 provides descriptive statistics on the variables used in this paper.

Insert Table 2 about here

Insert Table 3 about here

2.4 Indicator construction

For this paper, it is necessary to construct a large amount of citation indicators. Bakker et al. (2016) showed a variety of indicators that can be constructed and the vast differences between them. In this paper we will focus on the most widely used and applied indicators. We based these on the two most used data sources, the EPO and the USPTO as well as two patent family definitions: the previously mentioned DOCDB family and the larger INPADOC patent family, which is viewed as measuring larger innovations. We denote them by their main characteristic: EPO count counts the number of times an EPO application is cited by other EPO applications, USPTO count does the same for USPTO applications. DOCDB count counts the number times a DOCDB patent family is cited by other DOCDB patent families,

[5] The second highest level of aggregation of the International Patent Classification system, also known as the class level. In our data we observe 128 categories.
[6] For 7 patents no applicant was identified, these applications were not included in the analysis.

and INPADOC count does the same for INPADOC patent families. Their exact definitions are provided in Table 2.

3 Results

3.1 The effects of patent families and patent offices

With the advent of larger databases such as the PATSTAT database that is used in this paper, it became possible to construct patent citations based on patent families and that use citation data from more than 1 office. Bakker et al. (2016) found that citation indicators that are derived from these different measures vary greatly. It is therefore of great interest to understand how citations exactly reveal value. In this exercise, we will split this analysis first between citations from the same office as the focal application (i.e. the application for which we observe renewal) and citations from other offices.

Next we compute the difference between citations made to the focal application or to a DOCDB family member of the application, while also observing the effects of correcting for DOCDB family, i.e. preventing double counting citations when they come from the same family. Bakker et al. (2016) indicated that the INPADOC patent family indicator may be a good candidate for a generic citation indicator as it prevents large differences from occurring when different patent databases are used. Therefore we will extend this analysis to the INPADOC patent family. Because the INPADOC patent family is larger than the DOCDB family, we will calculate the revealed value from citations that are received by an INPADOC family member that is not a DOCDB patent family member. Again we also compute the effect of the family correction, but now for an INPADOC patent family correction. We computed all these measures for both renewal at EPO and renewal at USPTO. The results are listed in table 4.

Insert Table 4 about here

Table 4 lacks the analysis using citations to INPADOC family members to estimate renewal at USPTO. The reason for this lack is that the coefficients related to the INPADOC additions were sufficiently negative[7] to create a situation in which for at least some of the observations the non-linear term becomes imaginary. Since we assume that citations do not relate negatively with renewal, we see this as an indication that INPADOC added citations do not significantly add to the explanatory value of the citation indicator to explain patent renewal of USPTO patents.

From the other analyses presented in table 4, we can deduce the following. The difference in relative importance between citations from within the same patent system and citations from outside is present but minor. Citations to EPO patents from within the EPO system are more valuable, while citations to

[7] Different analyses with the coefficient restrained to a positive value revealed values very close to 0.

USPTO patents from within the USPTO system are less valuable. Differences are greater when considering citations to DOCDB family members. These citations are substantially less valuable than citations to the patent document itself, in both analyses of USPTO and EPO renewal. Citations to INPADOC family members are even less valuable, leading to the aforementioned problems of estimating them at USPTO. Finally, correcting for DOCDB and INPADOC patent families on the citing side is often counterproductive. The coefficients for these corrections are either negative or insignificant.

The analyses in table 4 show that not all citations are equal in terms of explanatory power. We therefore performed horse race regressions for the 4 basic indicators that are used in this paper: EPO count, USPTO count, DOCDB count and INPADOC count. We performed this analysis to give a better insight in the results of table 4 as well as to give researchers an idea about the efficacy of each indicator. In order to present a complete analysis we also reviewed how well the USPTO count indicator performs for the renewal of their EPO DOCDB family members and vice versa. There are some cases in which there are multiple granted patents of the same family in the same offices. Therefore these analyses have a few patents that are counted twice. This analysis is presented in table 5 for renewal of EPO patents and in table 6 for renewal of USPTO patents.

Insert Table 5 about here

Insert Table 6 about here

In these analyses we observe that there is not one indicator that performs best at explaining both the renewal of EPO patents and the renewal of USPTO patents. USPTO count does best for USPTO renewal, while INPADOC count performs best for EPO renewal. DOCDB count appears as a compromise, performing reasonable at explaining both EPO and USPTO renewal. The added value of citations to family members is much bigger for EPO renewal than it is for USPTO renewal. The cross analyses (USPTO count for EPO renewal and vice versa) show that the EPO count indicator is bad at explaining USPTO renewal, while the USPTO count indicator is reasonable at explaining EPO renewal, yet worse than the DOCDB count and INPADOC count indicators. These findings were also found to be robust when survival analyses instead of logit analyses were used.

3.2 Citations over time

A major drawback of using patent citations as an indicator is that citations are not given instantaneously but accumulate over time. When using patent citation a truncation problem is thus immediately presented as it is unknown whether a patent will receive a citation after the latest year recorded in the citation data. Researchers often resort to using citation windows in order to ensure truncation does not favor earlier patents, which have more time to be cited. Alternatively one can also try to estimate future citations as is done in the work of Trajtenberg(1990).

The use of citation windows brings with it the issue of estimating in what time period most information regarding the value of a patent is revealed by its forward citations. The approach that will be used in this paper relies on estimating the relative weight of citations from different time periods. The advantage from this approach is that not only it will reveal how citation windows operate, it will also provide a deeper insight into the underlying relation between citation timing and the value it reveals. This will in turn provide a better understanding of the relation between citations and value in general.

There are several theories that would correlate citation timing with value: The theory of preferential attachment in the patent citation system predicts that, ceteris paribus, patents that are already cited will be cited more in later periods. Therefore later citations will reveal less value, as the value of the patent was already revealed by the earlier citations. The works of (Hung and Wang 2010; Valverde et al. 2007) support the notion that the patent citation network displays signs of preferential attachment. Alternatively, later citations could be more informative because they were given at a time when the impact of the invention that a patent refers to is much better understood. Finally, later citations may indicate that the knowledge contained in the patent is relevant for a longer period.

We constructed indicators based on the year of the citing application using the relative citation year, which we define as the difference between the application year of the citing application and the application year of the cited application. This was also done for family based indicators, therefore it is possible that two applications from the same patent family have the different family citation counts if they were filed in different years. Finally we used adapted logistic regressions to determine the relative weights for the 4 basis indicators: EPO count, USPTO count, DOCDB count and INPADOC count for renewal at both the EPO and the USPTO. The EPO count and USPTO count weights are only computed for renewal at their respective offices. The results of this exercise are depicted in figure 1 for EPO renewal and figure 2 for USPTO renewal.

Insert Figure 1 about here

The indicators for EPO renewal all behave relatively similar: After a short decrease at the second relative citation year, they all display an increasing function with the EPO count having the strongest increase. Interestingly, the family indicators behave very similar despite being quite different in terms of descriptive statistics. This increasing importance of later citations provides credence to the theory that citations indicate a relevance in the peer community. A patent having citations long after its application indicates that its relevance persist over time and therefore that the patent is more likely to play a fundamental role in the innovation system. This is then reflected in the decision of the owner to keep renewing the patent.

Insert Figure 2 about here

3.3 Examiner, applicant and self-citations

It is common to view patent citations as a counterpart to citations to scientific articles. Patent citations are thus often viewed as representing knowledge flows (e.g. Jaffe et al. 1992, Jaffe and Trajtenberg 1999). Yet a substantial number of these citations are not added by the applicants of the invention but by patent examiners. In the application process, an applicant may first list relevant prior art, which will then be evaluated by the patent examiner. The examiner will also do his own research resulting in more prior art.

These examiner added citations lead to a situation in which patent inventors are often unaware of the prior art that is cited in their application(e.g. Jaffe et al. 1998,2000). Therefore patent citations do not necessarily always represent direct knowledge flows. Patent citations delivered by applicants may still be more valuable through a different mechanism: in the application process, applicants introduce prior art before the examiner. Therefore the applicant may be able to select the most relevant prior art, leaving less relevant prior art for the examiner to include. However, patent examiners may have a more unbiased view of the patent system and therefore may relevant prior art from sources unknown to the applicant. Therefore citations added by examiners may hold value, which may even be higher than citations added by applicants.

Furthermore, citation practices vary across different patent offices. In particular the USPTO requires that applicants disclose all known prior art when they apply for a patent in what is known as the 'duty of candor'. Other large patent office's such as the EPO and JPO do not have such a requirement, even though they do accept citations to relevant prior art from applicants. This 'duty of candor' has led to a relatively large number of applicant based citations in the USPTO patent system, compared to that of the EPO. Therefore the USPTO system appears to be best suited to test the differences between examiner and applicant citations.

There exist also difference between citations that derive from patents that share one or more applicants with the cited application. Considering these "self-citations" there are two schools of thought: The first school is derived from the study of scientific citations, e.g. see Van Raan(1996), and believes that self-citations should be corrected for. On the other side of the spectrum, it is argued that self-citations do provide a valuable indicator of innovative quality. This is due to the fact that self-citations indicate that the owner (i.e. the applicant) pursued further activity in the area. Therefore the area is apparently promising for the applicant which may be due to the quality of the work that was done before and is represented by the cited patent. Thus patents with many self-citations could indicate building stones for large projects within the firm (Narin et al. 1987, Belenzon, 2012). It has already been shown that a count of self-citations could perform better at explaining patent renewal than a count of other citations by Thomas (1999). We include self-citations in this analysis to distinguish between the revealed value of the citations of the same applicant and those of other applicants. Additionally, we want to determine if self-citations reveal a different value when they are added by examiners or applicants.

We performed an analysis differentiating between examiner, applicant and self-citations using an adapted logit regression. However, we note that the USPTO splits the citations in three: citations from the applicant, citations inserted during the search part of the examination procedure and citations to yet unpublished patents. Since the last two types of citations are likely inserted by examiners, we grouped them together. We also note that citations can, in principle, be duplicated when examiners and applicants cite the same application. The results of this analysis are shown in Table 7.

Insert Table 7 about here

The results from table 7 suggest first of all that applicant citations reveal more value than examiner citations per citation. Therefore the results noted in most of the literature are confirmed: applicant citations are more informative than examiner citations. This result is however, in direct contradiction to the results presented by Hegde and Sampat (2008). This may very well be due to the different functional form that was used, or the different and larger sample utilized in this paper. Second, self-citations reveal more value regardless if they are given by examiners or applicants. This severely strengthens the argument that self-citations mostly reveal that a patent is part of an ongoing effort of the applicant. It is interesting that self-citations given by applicants represent even more value than those given by examiners. This may be due to applicants picking patents that are still relevant and therefore longer renewed.

Applicant, examiner and self-citations have a different added value, the literature indicates that this is because these citations reveal value through different processes. It is likely that these processes differ even more when we include a time dimension in our analysis. Differences in how the timing of the citation will yield additional insights in these processes. Therefore the absolute values of the weights should be interpreted with caution. The resulting coefficients and confidence intervals of these regressions are presented in figure 3.

Insert figure 3 about here

Figure 3 clearly shows that the relative value of a citation is heavily dependent on the relative citation year. The relative value of applicant and examiner based citations approaches the lower bound of 0 after a few years, while the relative value of self-citations steadily increases over time. The relative decrease in citation importance over time lends credence to the preferential attachment theorems that we noted, while the increase over time of the importance of self-citations gives more weight to the argument that self-citations represent an ongoing effort on the part of the firm.

3.4 Citations by type and reason

Prior art in patents is cited for several reasons, out of which the majority fall into one of the following categories: to challenge a claim on the novelty criterion, to challenge a claim on the inventive step criterion or to give additional information to help understand the patent better. Citations can be made to single patents or to groups of patents. It is likely that the reason why a patent is cited by subsequent patents is a factor in determining the value indication of the citations. The use of patent databases that list these reasons has allowed researchers to take these reasons into account (e.g. Von Graevenitz et al. 2011).

To analyze the relative importance of these citations, we once again employ the relative weight methodology. We used the EPO indicator as the EPO search reports are the best reported in the PATSTAT database. Moreover, the EPO does not have a large tradition of applicants introducing citations and we feel therefore that the EPO citations are probably less influenced by these inclusions.

Only citations that were listed by EPO examiners in their search reports are taken into consideration. The most important reasons patents are cited, as listed in the EPO, is that they are prior art that is relevant to disprove novelty (type XN), prior art that is relevant as a background (type A) or prior art that is relevant in a larger group of patents (type Y). Recently patents are also classified as being relevant to disprove inventive step (type XI), but these citations are rare in the PATSTAT 2013 database. Finally patents can also cited because the applicant brought them in (type D). These citation categories as well as several smaller categories are included in the analysis; see Table 2 for exact definitions.

Some citations fall in several categories. Moreover, due to the database considerations of the EPO citation database less important categories may sometimes be omitted in a process that is explained in the resources provided by the EPO(2013). Therefore patent applications may receive multiple citations, albeit in different categories, from the same patent application. We calculated the relative weights of these different citation categories using the relative weight analysis, the results of which can be found in Table 8.

Insert Table 8 about here

The results in table 8 indicate that the differences between XN, Y and A references are insignificant: An F-test confirms this. It appears there is too little information in our sample to meaningfully estimate the relative weight of the XI citations or any of the less frequently used categories as seen by the insignificant, but high coefficients of these categories. The D category(i.e. the citations added by applicants) is substantially bigger than any of the other categories mentioned. This confirms the previous finding that citations from applicants are more informative than citations from examiners.

Despite finding that overall citations attributed for different reasons have the same informative value, it may still be possible that the relative value over time of these citations follows different patterns. We therefore again estimate these curves using the same approach as explained in the previous subsection. The results of this exercise for the citation counts of type XN, Y and A can be found in figure 4.

Insert Figure 4 about here

From figure 4 we can determine that there is a slight trend towards higher weights for later citations in all categories. Yet because of the high standard errors, relative to the weights we can't conclude that the citation types have significant different patterns. This is different for the EPO type D counts (i.e. the applicant citations). The relation between these citations and relative value is depicted in figure 5. It was necessary to depict the weights of self-citations on a different axis as they were an order of magnitude higher than the weights of other categories. Interestingly, the relative weights have a strongly increasing weight over time, reaffirming the continuing project theory of Narin et al. (1989).

Insert Figure 5 about here

----

4 Conclusion

This paper has set out to provide a deeper insight in the relation between patent citations and patent value. Patent value was observed using patent renewal as it is the only disaggregated directly observed measure of the revealed value of a patent to its owner. In order to better understand the relation between citations and this observed value, an adapted logistic regression was designed to determine the relative weights of different types of citations. These weights can then be used to understand the relative contribution of these different types of citations in the value estimate of the patent.

The results with respect to patent families indicate that the relation between a patent application and its equivalents at other offices is more complicated than previously understood. Citations made directly to the patent reveal more value than citations to its family members. This may indicate that the value of a patent is determined by its role in the local innovation system rather than its role in the global innovation system. Another interesting find is that correcting for patent families on the citing side has low and sometimes opposite effects. Therefore patent family correction on this side is shown to be both small (Bakker et al. 2016) and irrelevant or even harmful. Finally, the 'horse-race' regressions with respect to family indicators reveal that the efficacy of the indicator depends on the value indicator used. The INPADOC count indicator performs best to explain EPO renewal, while the USPTO count indicator performs best to explain the renewal of USPTO patents. DOCDB corrected patent citations appear as a compromise, which explains both EPO and USPTO renewal close to the optimal .

The timing of patent citations was also found to be a crucial factor. For most indicators, later citations had a higher relative value than earlier citations. It is possible that this is due to later citations coming from actors that had a longer time to observe the relative value of a patent. Alternatively later citations indicate that the relevance of a patent persist over time. Later analyses also indicate that this effect is to some extend driven by self-citations, whose relative importance increases greatly over time.

Patent citations also, as is commonly referenced in the literature, differ in the way they originate: self-citations are more correlated with private value than from other applicants, which in turn are relatively more valuable than examiner citations. The contribution of this paper is that besides confirming the literature, it is shown that the time profiles of these citations is also structurally different. Examiner citations loose importance at a steady pace, while applicant citations at the USPTO have an inverted U shape. Applicant based citations at the EPO and self-citations only increase over time. This also highlights the differences between the EPO patent system and the USPTO system with respect to applicant citations. Interestingly, the reason why a citation is given is of remarkable little importance, with X,Y and A citations at the EPO not having a statistically significant difference between them.

It is clear that the relation between patent citations and value is complicated and dependent on many factors. This helps confirm the notion that patent citations reveal value through different processes, which are not always well understood. It is surprising that after decades of research into this indicator, as well as its continuous use that it is far from clear what exactly is the reason that citations correlate with value. Therefore this paper is to be considered an initial foray in the exact meaning of patent citations. It would be prudent to, in future research, relate other indicators of value to these different citation characteristics. In that way the relation between citations and revealed value can be better understood.

Overall, patent citations have been found to explain little of the variance in patent value. Even the most detailed deconstructions of the citation indicator in this paper gave only a few extra percent increases in the recall and precision of the analyses; a result that replicates the findings of several other large scale studies. There are several theories that explain why patent citations should correlate with value. Yet there are few (e.g. preferential attachment) that explain why patent citations correlate so poorly with value of any kind. Despite this lack of explanatory power, patent citations are still the most used indicator to assess this very quality. It may very well be that this is due to a lack of better measures. There is thus a necessity to find better measures of patent quality. However, at the moment we have to warn against any interpretation of patent citations as a well-founded proxy of innovative output.

The results of this paper indicate that it may be of interest to researchers that may which to split citations into different categories which are more relevant to their work. Additionally, researchers should seriously consider the idea of weighting citations. Furthermore, this paper highlights the importance in choosing a citation indicator that adequately reflects the research objective, a call that was made in Bakker et al. (2016) as well. Moreover, a better look should be given into other patent based indicators that use patent citations in one way or another. This pertains to Trajtenbergs et al. (1997) generality, Shane(2001) radicalness, as well as other radicalness indicators (e.g. Dahlin and Behrens,2006) and breakthrough indicators (Ahuja and Lampert,2001). Indicators that use backward citations may be vulnerable as well since citations apparently have a time dimension and may indicate ongoing projects. Again this would affect other indicators: Trajtenberg et al. (1997) originality measure and the new origins indicator of Verhoeven et al. (2016). Finally, the study of knowledge spillovers should also more explicitly factor in the time dimension. Finally, the results of this paper may also be relevant on the study of academic citations. The time effects, especially in regard to self-referencing, observed in the relevance of patent citations may very well also be present in other bibliographic research such as the study of academic citations.

In conclusion, the results of this paper show the relevance of the continued study of patent citations as new properties of citations have been discovered This paper shed new light on the relationship between patent value and patent citations using fairly new dimensions such as patent family correction and citation timing. With the wealth of information present in the currently available patent data more can - and should- be done to investigate what drives the relation between patent citations and value. Finally, researchers should be more mindful of the properties we presented and apply them when using patent citations as a proxy for patent value.

5 References

Ahuja, G., & Morris Lampert, C. (2001). Entrepreneurship in the large corporation: A longitudinal study of how established firms create breakthrough inventions. *Strategic Management Journal*, *22*(6-7), 521-543.

Albert, M. B., Avery, D., Narin, F., & McAllister, P. (1991). Direct validation of citation counts as indicators of industrially important patents. *Research policy*,*20*(3), 251-259.

Albrecht, M. A., Bosma, R., van Dinter, T., Ernst, J. L., van Ginkel, K., & Versloot-Spoelstra, F. (2010). Quality assurance in the EPO patent information resource. *World Patent Information*, *32*(4), 279-286.

Alcacer, J., & Gittelman, M. (2006). Patent citations as a measure of knowledge flows: The influence of examiner citations. *The Review of Economics and Statistics*, *88*(4), 774-779.

Allison, P. D. (1999). Comparing logit and probit coefficients across groups. *Sociological methods & research*, *28*(2), 186-208.

Arts, S., Appio, F. P., & Van Looy, B. (2013). Inventions shaping technological trajectories: do existing patent indicators provide a comprehensive picture?.*Scientometrics*, *97*(2), 397-419.

Bakker, J., Verhoeven, D., Zhang, L., & Van Looy, B. (2016). Patent citation indicators: One size fits all?. *Scientometrics*, *106*(1), 187-211.

Bakker, J. Van Looy, B (2015). A critical analysis of patent citations as a measure of patent value, paper presented at the DRUID conference, June 2015

Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks.*science*, *286*(5439), 509-512.

Belenzon, S. (2012). Cumulative Innovation and Market Value: Evidence from Patent Citations. *The Economic Journal*, *122*(559), 265-285.

Bessen, J. (2008). The value of US patents by owner and patent characteristics. *Research Policy*, *37*(5), 932-945.

Carpenter, M. P., Narin, F., & Woolf, P. (1981). Citation rates to technologically important patents. *World Patent Information*, *3*(4), 160-163.

Chen, Y. S., & Chang, K. C. (2010). The relationship between a firm's patent quality and its market value—the case of US pharmaceutical industry. *Technological Forecasting and Social Change*, 77(1), 20-33.

Criscuolo, P., & Verspagen, B. (2008). Does it matter where patent citations come from? Inventor vs. examiner citations in European patents. *Research Policy*, *37*(10), 1892-1908.

Dahlin, K. B., & Behrens, D. M. (2005). When is an invention really radical?: Defining and measuring technological radicalness. *Research Policy*, *34*(5), 717-737.

Dosi, G. (1982). Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change. Research policy, 11(3), 147-162.

EPO (2013). ANNEX XIV - Categories of Cited Documents in: *EXCHANGE FORMAT EPO - Patent Information Resource.* distribution by : Exchange Partners

EPO(2015). Schedule of fees available at: http://www.epoline.org/portal/portal/default/epoline.Scheduleoffees-available, last visited on 29-06-2015

Gambardella, A., Harhoff, D., & Verspagen, B. (2008). The value of European patents. *European Management Review*, *5*(2), 69-84.

Gay, C., & Le Bas, C. (2005). Uses without too many abuses of patent citations or the simple economics of patent citations as a measure of value and flows of knowledge*. *Economics of Innovation and New Technology*, *14*(5), 333-338.

Griliches, Z. (1998). Patent statistics as economic indicators: a survey. In *R&D and productivity: the econometric evidence* (pp. 287-343). University of Chicago Press.

Gittelman, M. (2008). A note on the value of patents as indicators of innovation: Implications for management research. *The Academy of Management Perspectives*, *22*(3), 21-27.

Hall, B. H., Jaffe, A., & Trajtenberg, M. (2005). Market value and patent citations. *RAND Journal of economics*, 16-38.

Hall, B. H., Thoma, G., & Torrisi, S. (2007). The market value of patents and R&D: evidence from European firms. In *Academy of Management Proceedings* (Vol. 2007, No. 1, pp. 1-6). Academy of Management.

Harhoff, D., Narin, F., Scherer, F. M., & Vopel, K. (1999). Citation frequency and the value of patented inventions. *Review of Economics and statistics*,*81*(3), 511-515.

Harhoff, D., Scherer, F. M., & Vopel, K. (2003). Citations, family size, opposition and the value of patent rights. Research Policy, 32(8), 1343-1363.

Hegde, D., & Sampat, B. (2009). Examiner citations, applicant citations, and the private value of patents. *Economics Letters*, *105*(3), 287-289.

Hung, S. W., & Wang, A. P. (2010). Examining the small world phenomenon in the patent citation network: a case study of the radio frequency identification (RFID) network. *Scientometrics*, *82*(1), 121-134.

Jaffe, A. B., Trajtenberg, M., & Henderson, R. (1992). *Geographic localization of knowledge spillovers as evidenced by patent citations* (No. w3993). National Bureau of Economic Research.

Jaffe, A. B., Fogarty, M. S., & Banks, B. A. (1998). Evidence from patents and patent citations on the impact of NASA and other federal labs on commercial innovation. *The Journal of Industrial Economics*, *46*(2), 183-205.

Jaffe, A. B., & Trajtenberg, M. (1999). International knowledge flows: evidence from patent citations. *Economics of Innovation and New Technology*, *8*(1-2), 105-136.

Jaffe, A. B., Trajtenberg, M., & Fogarty, M. S. (2000). *The meaning of patent citations: report on the NBER/Case-Western Reserve survey of patentees* (No. w7631). National bureau of economic research.

Kitch, E. W. (1977). The nature and function of the patent system. *Journal of law and economics*, 265-290.

Magerman, T., Van Looy, B., & Song, X. (2006). *Data production methods for harmonized patent indicators: Patentee Name Harmonization*. EUROSTAT Working Paper and Studies, Luxembourg.

Marchetti, C. (1980). Society as a learning system: discovery, invention, and innovation cycles revisited. *Technological forecasting and social change*, *18*(4), 267-282.

Maurseth, P. B. (2005). Lovely but dangerous: The impact of patent citations on patent renewal.*Economics of Innovation and New Technology*,*14*(5), 351-374.

Mowery, D. C., Nelson, R. R., Sampat, B. N., & Ziedonis, A. A. (2001). The growth of patenting and licensing by US universities: an assessment of the effects of the Bayh–Dole act of 1980. *Research policy*, *30*(1), 99-119.

Narin, F., Noma, E., & Perry, R. (1987). Patents as indicators of corporate technological strength. *Research policy*, *16*(2), 143-155.\

Pakes, A. (1986). *Patents as options: Some estimates of the value of holding European patent stocks* (No. w1340). National Bureau of Economic Research.

Peeters, B., Song, X., Callaert, J., Grouwels, J., & Van Looy, B. (2010). Harmonizing harmonized patentee names: an exploratory assessment of top patentees. *Eurostat Working Paper*.

Schankerman, M., & Pakes, A. (1985). *Estimates of the Value of Patent Rights in European Countries During thePost-1950 Period* (No. w1650). National Bureau of Economic Research.

Scherer, F. M., & Harhoff, D. (2000). Technology policy for a world of skew-distributed outcomes. Research Policy, 29(4), 559-566.

Shane, S. (2001). Technology regimes and new firm formation. Management science, 47(9), 1173-1190.

Squicciarini, M., Dernis, H., & Criscuolo, C. (2013). *Measuring patent quality: Indicators of technological and economic value* (No. 2013/3). OECD Publishing.

Thomas, P. (1999). The effect of technological impact upon patent renewal decisions. *Technology Analysis & Strategic Management*, *11*(2), 181-197.

Trajtenberg, M. (1990). A penny for your quotes: patent citations and the value of innovations. *The Rand Journal of Economics*, 172-187.

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

USPTO (2015), USPTO Fee Schedule, available at http://www.uspto.gov/learning-and-resources/fees-and-payment/uspto-fee-schedule -last visited 29-06-2015

Valverde, S., Solé, R. V., Bedau, M. A., & Packard, N. (2007). Topology and evolution of technology innovation networks. *Physical Review E*, *76*(5), 056118.

Van Raan, A. F. (1996). Advanced bibliometric methods as quantitative core of peer review based evaluation and foresight exercises. *Scientometrics*, *36*(3), 397-420.

Van Zeebroeck, N. (2011). The puzzle of patent value indicators. *Economics of Innovation and New Technology*, *20*(1), 33-62.

Verhoeven, D., Bakker, J., & Veugelers, R. (2016). Measuring technological novelty with patent-based indicators. *Research Policy*, *45*(3), 707-723.

Von Graevenitz, G., Wagner, S., & Harhoff, D. (2011). How to measure patent thickets—A novel approach. *Economics Letters*, *111*(1), 6-9.

Webb, C., Dernis, H., Harhoff, D., & Hoisl, K. (2005). Analysing European and International Patent Citations. *OECD Science, Technology and Industry Working Papers*, 2005/09, OECD Publishing.

Appendix A: EPO renewal

This paper is one of the few to employ uses patent renewal at the EPO as a dependent variable. It is therefore important to set out the patent application and maintenance regime at the EPO as well as to provide descriptive statistics to provide a better insight in this indicator.

Whenever a patent is applied for in the EPO system. It first enters an application phase at the EPO. In this phase the applicant provides a patent application and designates the national offices of the EPO(e.g. Germany and France) in which he wants the patent to applied in. During this phase maintenance fees need to be paid at the EPO. When the patent is granted, the patent enters a 'national' phase. In order to maintain the patent, owners need to pay maintenance fees to national offices. Therefore, the costs of maintaining a patent is from then depends on which national offices the owner chooses to maintain the patent at.

This procedure creates a situation in which renewal at the EPO is more complicated than renewal at the USPTO. We opted for the use of the Single Renewal Approach (SRA) described in van Zeebroeck(2011). This approach has the benefit that it produces a metric that is comparable with the USPTO renewal and avoids data issue due to missing data from several smaller patent offices that subscribe to the EPO, if we assume that EPO patents are unlikely to be renewed at a small office, while abandoned at larger better documented offices. In figure A1, we listed the offices of which patents that were maintained until their maximum lifetime at only 1 national office of the EPO. This table shows that the vast majority of these patents were maintained at large offices, thus providing support for this assumption.

Insert table A1

However, next to considering the individual value of the patent, we should also make note of the strategic benefits that accrue with holding the patent in several countries. These benefits arise from the integrated European markets. Having a patent in one European country blocks both production and distribution of infringing products in this country. Therefore countries that represent strategic markets are more interesting to maintain patents in than countries that are not. Moreover, by securing key markets it can become unprofitable for competitors to invest in infringing technology due to scale considerations. Thus a company may achieve an optimal result with maximum protection and lower maintenance costs by selectively abandoning patents in less strategically relevant offices. Using the SRA avoids the issues that arise from these strategic interactions: when a patent is abandoned in all offices, there are no strategic interactions. It is therefore likely that the value of the patent to the firm has slipped under any maintenance cost of any eligible national office at the EPO.

Finally, the SRA method also can be interpreted using a more theoretical consideration. When considering a single patent office, renewal can be thought of as a value floor: the patent is at least worth the maintenance costs paid. If a patent is allowed to lapse it is worth less than the maintenance cost (assuming full rationality and no financing constraints of the owner). When a patent is maintained at multiple offices, its value is at least the combined costs of its maintenance. If it is allowed to lapse it is worth less than the minimum costs of its maintenance. Hence, it is very hard to directly compare the value of patents maintained at multiple offices with those maintained at a single office as renewal decisions only signal a minimum value: patents renewed at a single office may be worth less or more than patents renewed at multiple offices. Yet comparing a the value of a patent not renewed at any office and a patent renewed at any other office is far simpler, patents that are renewed are worth more than patents that are not renewed. Implementing the SRA method, implies determining the difference between renewed and not renewed patents, while staying agnostic about the differences between patents renewed at multiple offices and patents renewed at a single office. Therefore in the complicated matter of EPO, the SRA method has also has theoretical advantages.

6 Tables

Table 1: An overview of selected studies that examine patent citations

Study	source	Independent Variable	Dependent Variable	Controls	Statistical Method	Number of observations	Result with respect to patent citations
Carpenter et al. (1981)	USPTO	Count	R&D award given	Year	2-way ANOVA	202	31.38 F score
Narin and Noma(1987)	USPTO	Count	Financial performance	N/A	Correlation	17	0.628 correlation
Trajtenberg(1990)	USPTO	Aggregated citations	Innovative value on industry level	N/A	Correlation	10	0.685 correlation
Albert et al.(1991)	USPTO	Groups	Relative importance in industry	Year/ company/ technology	ANOVA	77 in 8 groups	2.01 F score (only most cited group significantly different)
Harhoff et al. (1999)	USPTO	Log(count)	Replacement value of patent	none	OLS	192	6.3% of variation explained
Thomas (1999)	USPTO	Count	Renewal decision	ÌPC4, applicant type	Various non parametric analyses	189,359	Significant correlation between patent citations and renewal
Hall et al. (2005)	USPTO	Stock	Tobins Q	R&D, sales, year industry	Non-linear model	12188 (1983 with patents)	3.2% of variation explained
Hall et al. (2007)	EPO	Stock	Tobins Q	R&D, sales, year, industry	Market eq. estimation	1 779	Insignificant
Bessen (2008)	USPTO	Count	Renewal value	Applicant, backward citations. Generality Originality	OLS	48990	4-7% of variation explained
Gambardella et al. (2008)	EPO	Log(count)	Replacement value of patent	Year/country/ tech class[8]	OLS	8 217	1.4% of variation explained
Chen and Chang(2010)	USPTO	Average citations/pate nt	Market value of pharmaceutic al company	Sales, sales growth, other patent portfolio characteristics	Fixed effects OLS	370	Citation indicators are highly significant

[8] 30 classes of the ISI-INPI-OST classification

Arts et al.(2013)	USPTO	Groups of highly cited	Importance of invention	Year/USPC class	Logistic analyses	74 072	Citation indicators are highly significant

Table 2: Variables used in this paper and their definitions. *indicates categorical variables that may have multiple categories per patent application. In our analysis we applied a partial count metric when this occurred. Categories with less than 50 applications where grouped in a 'other' category. # indicates a continuous variable pertaining to the characteristics of the applicants of the application. When multiple applicants were present on the application the largest value was taken.

Indicator	Definition
Patent citation indicators	
EPO count	Count of citations that are received by EPO applications from other EPO applications
USPTO count	Count of citations that are received by USPTO applications from other USPTO applications
DOCDB count	Count of citations that are received by the application or members from its DOCDB family from other DOCDB families
INPADOC count	Count of citations that are received by the application or members from its INPADOC family from other INPADOC families
Control variables	
Nr. Countries	Number of distinct patent offices in which the DOCDB family of the patent has at least 1 application present .
IPC3*	Variable to indicate if the IPC3 class (e.g. A01) is present in the patent application
Year	Year in which the application was applied for at the patent office
Ln(Applt. Size) #	Logarithm of the total number of patents filed by applicant
Applicant type*	Type of applicant: Company, government, hospital, individual, university or unknown
Applicant experience#	Years between filing of current patent and that of the first application filed by the applicant
Applicant country*	Country in which the applicant resided at time of filing the patent
Co-patented	Dummy to indicate if the patent has more than 1 applicant
Exclusive patent citation indicators	
EPO out	Count of citations made to an EPO patent by non-EPO patents
USPTO out	Count of citations made to an USPTO patent by non-USPTO patents
DOCDB EPO	Count of citations made to a DOCDB family member of an EPO patent
DOCDB USPTO	Count of citations made to a DOCDB family member of an USPTO patent
DOCDB cor.	Correction made by deduplication of citations made by one DOCDB patent family to another.
INPADOC add.	Count of citations made to INPADOC, but not DOCDB, family members of the focal patent
INPADOC cor.	Correction made by deduplication of citations made by one INPADOC patent family to another.
Self-citation indicators	
DOCDB(self)	Count of citations that are received by the application or members from its DOCDB family from other DOCDB families which have the same applicant
DOCDB(not self)	Count of citations that are received by the application or members from its DOCDB family from other DOCDB families which do not have the same applicant
Citation by origin and reason	
USPTO(APP)	Count of citations that are given by applicants of USPTO patents to the USPTO application
USPTO(EXA)	Count of citations that are given by examiners of USPTO patents (in their search report) to the USPTO application
EPO type A	Count of awarded citations made to understand the state of the art in the field
EPO type D	Count of citations made by the applicant
EPO type E	Citations patents that have earlier priority dates but later publication dates
EPO type T	Citations made to reference a principle or theory underlying the invention
EPO type XI	Citations that are made to prejudice inventive step
EPO type XN	Citations that are made to prejudice novelty
EPO type Y	Citations to patents that need to be seen as part of a larger group with other patents that are cited under the same code
EPO type L	Citations for other reasons than those listed in the previous EPO indicators
Citation window analysis	

DOCDB year *	Count of citations added to the DOCDB indicator in the *th year after the first application was filed for the DOCDB family
DOCDB * window	Count of citations taking only into account the first * years after the first application was filed for the DOCDB family
DOCDB * window (out)	Count of citations that are excluded by taking a * citation window

Table 3: Descriptive statistics of variables used in this paper

Variable	N	Mean	Std. Dev.	Min	Max
Full term EPO	338405	0.21	0.40	0	1
Full term USPTO	560122	0.52	0.50	0	1
Nr. Countries	571814	7.11	4.21	2	51
Ln(Applt. Size)	571814	7.47	3.33	0	12.9937
Applicant experience	571814	0.06	0.24	0	1
Co-patented	571814	36.37	30.86	0	146
EPO count	571814	1.85	3.93	0	311
USPTO count	571814	16.17	30.33	0	2802
DOCDB count	571814	26.53	60.55	0	3146
INPADOC count	571814	32.82	114.00	0	4747
EPO out	571814	3.33	7.92	0	548
USPTO out	571814	3.23	10.89	0	2267
DOCDB EPO	571814	34.99	97.14	0	5217
DOCDB USPTO	571814	20.76	83.76	0	4121
DOCDB cor.	571814	-13.63	40.24	-2080	0
INPADOC add.	571814	12.64	136.98	-832	6945
INPADOC cor.	571814	-19.98	86.21	-3240	0
DOCDB(self)	571814	3.67	10.85	0	481
DOCDB(not self)	571814	22.86	55.22	0	3034
DOCDB 5 window	571814	7.65	16.08	0	588
DOCDB 5 window(out)	571814	32.51	87.43	0	5096
DOCDB 5 window	571814	30.36	75.67	0	2637
DOCDB 10 window(out)	571814	9.80	28.54	0	2589

Table 4: Adapted logistic regressions related to different sources of citations. The dependent variable: patent being renewed to its maximum term at the EPO or USPTO

	(1) Full term EPO	(2) Full term EPO	(3) Full term EPO	(4) Full term USPTO
Linear				
Nr. Countries	0.0635***	0.0513***	0.0521***	0.0250***
	(0.00136)	(0.00140)	(0.00140)	(0.000856)
Ln(applt. size)	0.00792**	0.00342	0.00275	0.0398***
	(0.00274)	(0.00277)	(0.00277)	(0.00164)
Applicant experience	0.0301	0.0192	0.0236	-0.0465*
	(0.0361)	(0.0371)	(0.0371)	(0.0224)
Co-patented	-0.00296***	-0.00174***	-0.00160***	-0.00428***
	(0.000260)	(0.000263)	(0.000263)	(0.000148)
Constant	-24.14	-25.53	-25.08	-24.57
	(1328.1)	(1558.9)	(1542.9)	(813.4)
Non-linear				
EPO count	0.764***	0.611***	0.964***	
	(0.0566)	(0.0549)	(0.0888)	
EPO out	0.597***	0.252***	0.504***	
	(0.0389)	(0.0321)	(0.0542)	
DOCDB EPO		0.173***	0.348***	
		(0.0151)	(0.0310)	
USPTO count				0.617***
				(0.0233)
USPTO out				0.831***
				(0.0389)
DOCDB USPTO				
DOCDB cor.		-0.247***		
		(0.0228)		
INPADOC add.			0.129***	
			(0.0203)	
INPADOC cor.			0.0500	
			(0.0312)	
IPC3 Dummies	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes
sector dummies	Yes	Yes	Yes	Yes
Country dummies	Yes	Yes	Yes	Yes
N	292625	292625	292625	519298
Pseudo R^2	0.045	0.067	0.067	0.080
Log Likelihood	-135858	-132652	-132635	-330947
Recall(%)	1.86	4.43	4.30	67.60
Specificity(%)	99.67	99.34	99.35	59.18

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

Table 5: Horse-race logistic regressions to determine which citation indicator best explains EPO renewal

	(1) Full term EPO	(2) Full term EPO	(3) Full term EPO	(4) Full term EPO	(5) Full term EPO
Ln(1+EPO count)		0.433*** (0.00631)			
Ln(1+USPTO count)			0.374*** (0.00462)		
Ln(1+DOCDB count)				0.601*** (0.00614)	
Ln(1+INPADOC count)					0.593*** (0.00599)
Nr. countries	0.0647*** (0.00135)	0.0631*** (0.00137)	0.0701*** (0.00119)	0.0532*** (0.00140)	0.0527*** (0.00140)
Ln(applt. size)	0.0161*** (0.00272)	0.00607* (0.00275)	0.0126*** (0.00252)	0.00351 (0.00277)	0.00308 (0.00277)
Applicant experience	0.0260 (0.0360)	0.0292 (0.0361)	0.0102 (0.0323)	0.0228 (0.0372)	0.0266 (0.0371)
Co-patented	-0.00306*** (0.000258)	-0.00298*** (0.000261)	-0.00333*** (0.000239)	-0.00165*** (0.000263)	-0.00133*** (0.000263)
Constant	-22.26 (843.3)	-24.80 (1708.8)	-25.58 (1500.9)	-24.13 (910.5)	-24.16 (916.1)
IPC3 Dummies	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes
sector dummies	Yes	Yes	Yes	Yes	Yes
Country dummies	Yes	Yes	Yes	Yes	Yes
N	292625	292625	330478	292625	292625
Pseudo R^2	0.027	0.044	0.060	0.063	0.064
Log likelihood	-138343.3	-135986.4	-158992.5	-133245.4	-133100.4
Recall(%)	0.819	2.090	5.508	4.141	4.644
Specificity (%)	99.84	99.62	99.03	99.37	99.24

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

Table 6: Horse-race logistic regressions to determine which citation indicator best explains USPTO renewal

	(1) Full term USPTO	(2) Full term USPTO	(3) Full term USPTO	(4) Full term USPTO	(5) Full term USPTO
Ln(1+EPO count)		0.281*** (0.00398)			
Ln(1+USPTO count)			0.407*** (0.00318)		
Ln(1+DOCDB count)				0.431*** (0.00357)	
Ln(1+INPADOC count)					0.409*** (0.00348)
Nr. countries	0.0287*** (0.000840)	0.0314*** (0.000802)	0.0267*** (0.000855)	0.0186*** (0.000857)	0.0190*** (0.000857)
Ln(applt. size)	0.0456*** (0.00161)	0.0406*** (0.00157)	0.0367*** (0.00164)	0.0369*** (0.00163)	0.0373*** (0.00163)
Applicant experience	-0.0420 (0.0221)	-0.0129 (0.0208)	-0.0402 (0.0225)	-0.0503* (0.0224)	-0.0529* (0.0224)
Co-patented	-0.00498*** (0.000145)	-0.00498*** (0.000142)	-0.00415*** (0.000148)	-0.00405*** (0.000147)	-0.00395*** (0.000147)
Constant	-26.23 (1674.5)	-23.92 (927.2)	-25.51 (1047.4)	-25.46 (1044.2)	-25.42 (1046.6)
IPC3 Dummies	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes
Sector dummies	Yes	Yes	Yes	Yes	Yes
Country dummies	Yes	Yes	Yes	Yes	Yes
N	519298	550857	519298	519298	519298
Pseudo R^2	0.054	0.062	0.078	0.075	0.074
Log likelihood	-340198.8	-357438.8	-331573.3	-332577.2	-332936.9
Recall(%)	67.52	69.55	67.50	67.26	67.18
Specificity(%)	55.79	54.70	59.05	59.04	59.21

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

Table 7: The relative weights of examiner(EXA) and applicant citations at the USPTO. Estimated using an adapted logistic regression, with USPTO renewal as a dependent variable.

	(1) Full term	(2) Full term
Linear		
Nr. Countries	0.0264*** (0.000855)	0.0242*** (0.000858)

Log (applt. size)	0.0374***	0.0256***
	(0.00164)	(0.00166)
Copatented	-0.0425	-0.0232
	(0.0224)	(0.0225)
Firm experience	-0.00418***	-0.00429***
	(0.000147)	(0.000148)
Constant	-27.44	-24.84
	(2040.0)	(1208.9)
Non-linear		
USPTO count (APP)	0.829***	
	(0.0297)	
USPTO count (EXA)	0.629***	
	(0.0256)	
USPTO count (APP)		0.613***
Non self –citation		(0.0225)
USPTO count (EXA)		0.335***
Non self –citation		(0.0166)
USPTO count (APP)		3.905***
self –citation		(0.210)
USPTO count (EXA)		2.720***
self –citation		(0.106)
IPC3 Dummies	Yes	Yes
Year dummies	Yes	Yes
Sector Dummies	Yes	Yes
Country dummies	Yes	Yes
N	519298	519298
Pseudo R^2	.078	0.081
Log likelihood	-331510	-330259
recall(%)	67.31	66.76
specificity(%)	59.28	60.52

Table 8: Adapted logistic regression to determine the relative value of different types of EPO citations in explaining EPO renewal.

	(1) Full term
Linear	
Nr. Countries	0.0627***
	(0.00137)
Log(applt. size)	0.00995***
	(0.00275)
Co-patented	0.0265
	(0.0361)
Firm experience	-0.00309***
	(0.000260)
Constant	-23.56
	(1246.3)
Non-linear	
EPO (type A)	0.511***
	(0.0330)
EPO (type D)	4.370***
	(0.228)
EPO (type XN)	0.550***
	(0.0537)
EPO (type XI)	1.167
	(0.789)
EPO (type T)	-0.336
	(0.290)
EPO (type L)	-0.167
	(0.751)
EPO (type E)	0.272
	(0.162)
EPO (type Y)	0.653***
	(0.0760)
IPC3 Dummies	Yes
Year dummies	Yes
Sector dummies	Yes
Country dummies	Yes
N	292625
Pseudo R^2	0.0428
log likelihood	-136107
recall(%)	1.907
Specificity(%)	99.62

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

Table A1: The office in which applications were maintained, if they were renewed to their maximum term at an EPO office.

office	Nr. applications	cum. Sum	cum. % of relative to the total number of fully renewed applications
GR	10	10	0.01%
DK	12	22	0.03%
AT	52	74	0.11%
ES	97	171	0.24%
BE	103	274	0.39%
SE	148	422	0.60%
CH	233	655	0.94%
NL	268	923	1.32%
IT	799	1722	2.47%
FR	1521	3243	4.64%
GB	2022	5265	7.54%
DE	4204	9469	13.56%

7 Figures

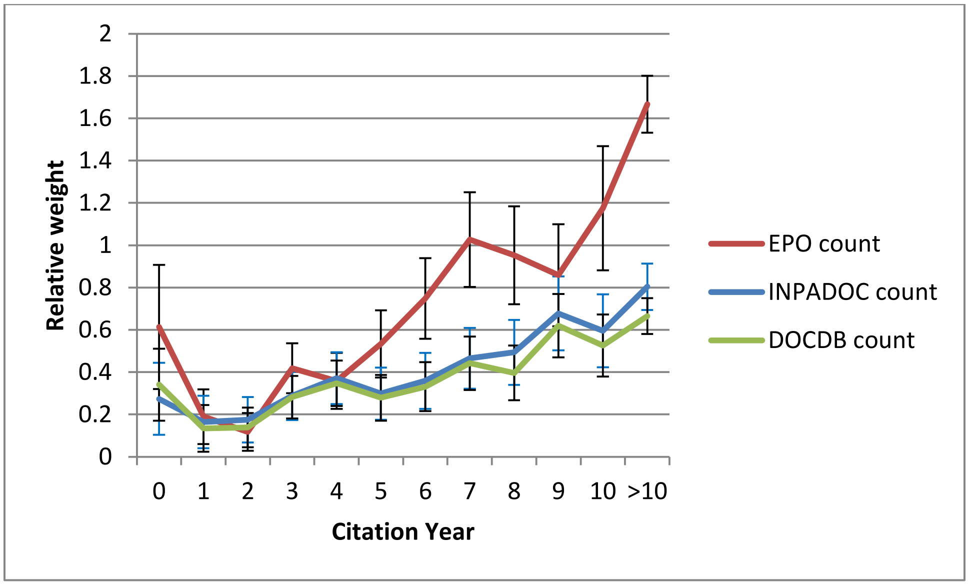

Figure 1:relative weights for different citation indicators that explain EPO renewal. Relative weights were calculated separately for each indicator.

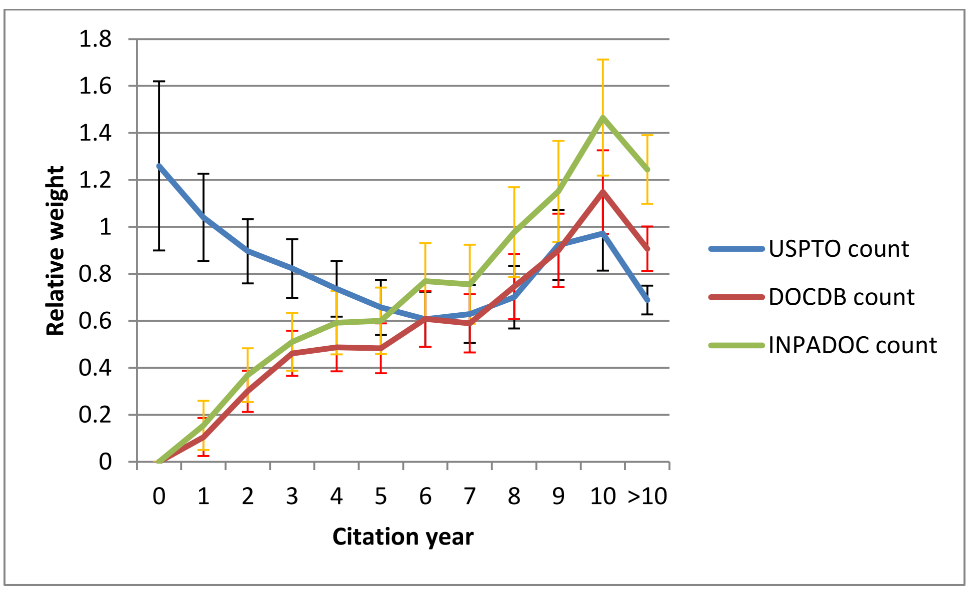

Figure 2 relative weights for different citation indicators that explain EPO renewal. Relative weights were calculated separately for each indicator.

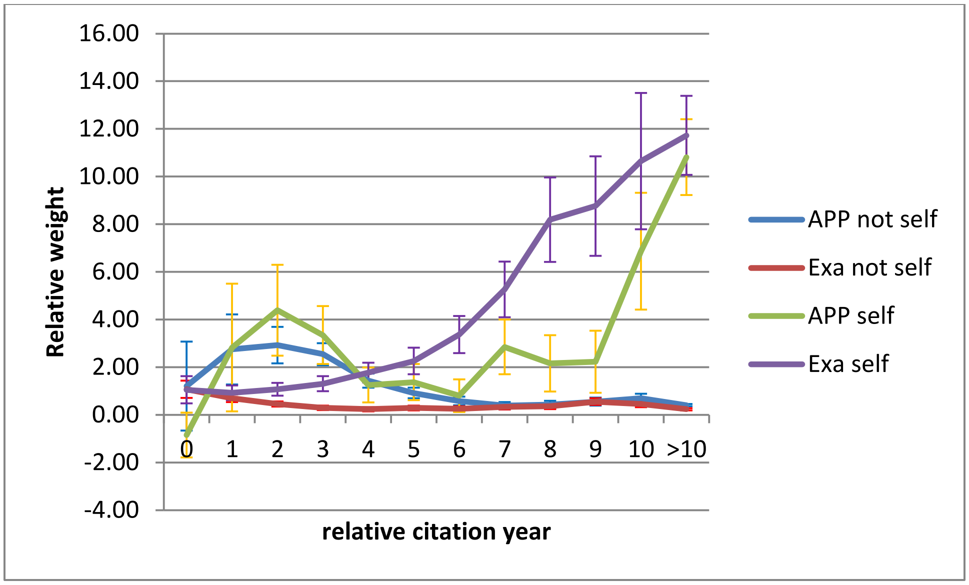

Figure 3:relative citation weights over time for different USPTO indicators that explain renewal at the USPTO. Error bars signal a 95% confidence interval.

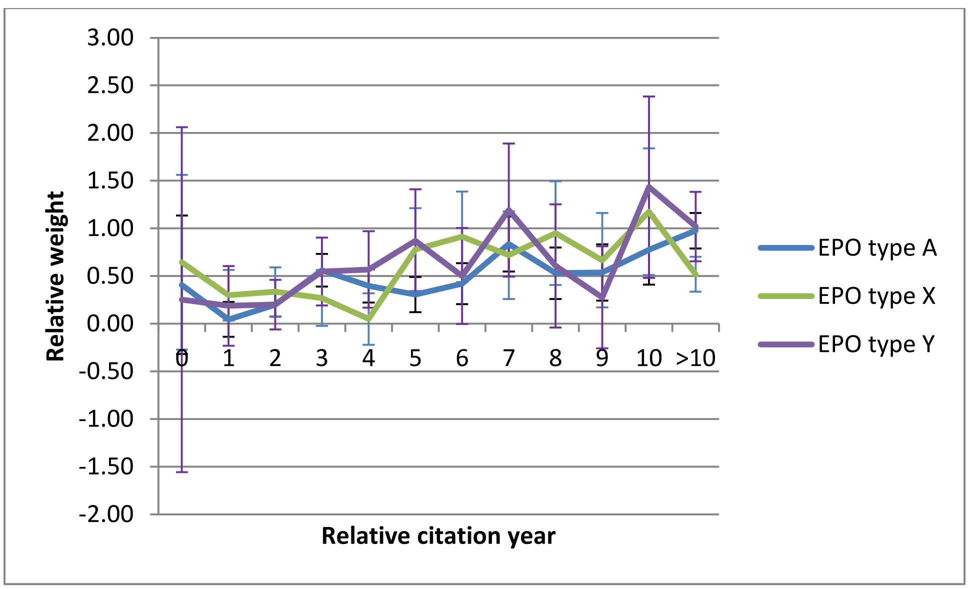

Figure 4: Relative weights for different types of EPO citation for different relative citation years.

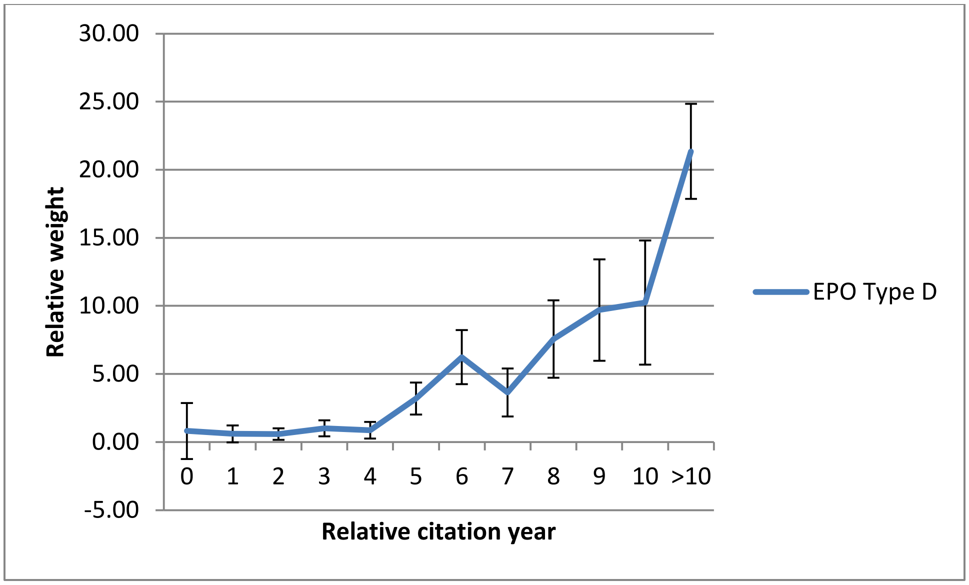

Figure 5: Relative weights for the EPO type D(i.e. applicant citations), for different citation years to explain renewal of EPO applications. Error bars signal a 95% confidence interval.