Can Patent Family Size and Composition Signal Patent Value?†

Francois Kabore,
Center for Research and Action for Peace (CERAP),
Abidjan, Ivory Coast
fpk2@georgetown.edu/francois.kabore@cerap-inades.org

Walter G. Park∗
American University,
Department of Economics,
Washington, D.C., USA
wgp@american.edu

September 2016

Abstract

Recent research has proposed a method of patent valuation based on weighting patent family size by the market size of the countries in the family. The premise is that inventors tend to seek greater international coverage for their more valuable patents. The paper presents a novel way to test the ability of market size-weighted patent families to predict patent value, and compares the method against extant measures of patent valuation based on patent citations and renewal behavior. We use forecasting techniques to show that the weighted patent family size measure outperforms other methods in terms of predicting patent life and the number of citations. An advantage of the weighted patent family size measure is that it is based on ex ante information.

Keywords: Patent valuation, Market Size, Patent Families, Citations, Renewal

JEL classification: O34, O33, F23, K11

†We thank Jonas Anderson, Robert Feinberg, Peter Hingley, Keun Lee, Alan Marco, Kara Reynolds, Wonkyu Shin, and seminar participants at STEPI and the U.S. Patent and Trademark Office (USPTO) for feedback and discussions. We also thank the Institute for Humane Studies for financial support.

*Corresponding author.
I. INTRODUCTION

Patent valuation is important to firms’ decision-making and strategies. For example, it shapes incentives for investments in innovation. It also plays a key role in licensing negotiations and in determining the relative bargaining strength of parties (Smith and Parr, 2000). Patent litigation is also affected by perceptions of patent value. The more valuable the patent rights are, the more committed patent holders will be to monitor and enforce their rights. Patent valuation is also important to policy-makers for assessing the impact of legal provisions or innovation policies on the level and quality of new technologies. Is the policy regime merely encouraging the filing of patents for low-valued inventions or generating transformative wealth-creating activities? The valuation of patents can also provide useful indicators for comparing the production of innovations across industries and countries.

However, the value of patents is not easily observed. Patent rights are not frequently or widely traded like financial instruments, such as equities. Large, international exchange institutions for providing markets for patent rights are absent. Few firms publicly disclose details about their patent transactions or technology transfer agreements. Furthermore, not all patent rights have a steady cash flow, since some (if not many) patented inventions are not commercialized, or are not “worked” (put into practice). Other problems may be that some patents are invalid (not correctly granted and will be invalidated upon challenge) or that asymmetric information exists (where the seller knows the quality of the underlying invention but the buyer does not) so that contract prices may distort the true valuation of patents.

Consequently, from a research standpoint, patent valuation is largely based on observing the characteristics of a patent or the behavior of a patent holder that may reveal information about the underlying value. For example, previous work by Bessen (2008), Lanjouw et al. (1998), Pakes and Simpson (1989), and Schankerman and Pakes (1986) examined patent renewal behavior. Given the

---

1 See Bessen and Meurer (2009) and Maskus (2006).
2 However, patent rights are tradable at auctions, held for example by Ocean Tomo, an intellectual property merchant bank. Fisher and Leidinger (2014), Odasso et al. (2015), and Mauck and Pruitt (2016) study patent values from Ocean Tomo auctions.
3 See Murphy et al. (2012) for a discussion of the impact of adverse selection on patent values.
costs of maintaining patent rights, the relatively most valuable patents will be selected by patent holders for renewal. Another approach has been to factor in a patent’s forward citations (see Trajtenberg (1990), Harhoff et al. (1999), Jaffe and Trajtenberg (2002), and Hall et al. (2005)). The more valuable a patent is the more likely it will find use in follow-on research or production and therefore be more often cited. Barney (2002) finds that citations and renewal are related in that patents that receive no forward citation in their first four years are less likely to be renewed than those with multiple citations. On the other hand, patent renewals and citations may not always yield the same ‘signal’. A patent may lapse (for lack of market value or other) and yet continued to be cited.

Another approach to gauging patent value is to examine patent family size (see Harhoff et al. 2003; Johnstone et al., 2012; Lanjouw and Schankerman, 2004; and Putnam, 1996). Patent family size refers to the number of countries in which an invention is protected by a patent; that is, the priority filing and the subsequent filings that emanate from it. Given the cost of acquiring patents in multiple jurisdictions, rights holders would reserve their more valuable technologies for international patenting. A variant of this approach is to consider triadic patents, where an invention is protected in the three major markets, namely the U.S., Japan, and the European Patent Office (EPO) (see Dernis et al, 2001), or transnational patents, which are patent families with at least one EPO or Patent Cooperation Treaty (PCT) filing.

---

4Patent citations are also used to trace knowledge diffusion (Hu and Jaffe, 2003).

5Moreover, renewal decisions may largely reflect private value while citations reflect some social value as well, considering the impacts of the patented technology on other inventions.

6There is also an extended patent family definition, as described in Hingley and Park (2003), where “a patent family encompasses all the documents related to the patents emanating from an invention, including documents that may cross relate to other inventions as well.” See also Martinez (2011) for a thorough survey of patent family definitions and methodologies.

7There are other potential indicators of patent value which we have not reviewed and critiqued, such as whether a patent has faced an opposition challenge (Harhoff and Reitzig, 2004), the number of claims in a patent (Tong and Frame, 1994), joint ownership of a patent (Briggs and Wade, 2014), auction prices (Nair et al., 2011; Sneed and Johnson, 2009), and filings strategies (van Zeebroeck and van Pottelsburg de la Potterie, 2011). Other approaches to determining patent value are event studies around court decisions or patent announcements (see Henry, 2013 and Austin, 1993). In this paper, we focus primarily on comparing our method of valuation to the most widely used indicators of patent value thus far, namely patent renewals and citations. Other indicators (like claims or ownership) are also ex ante measures and could in future be tested.
A key limitation of the citation and renewal methods is that patent values are assessed ex-post. As pointed out by Hall et al. (2005), a sufficient time is needed after a patent is granted to accumulate information about its citation. Furthermore, patent citations are not distinguished by how much impact the cited patents have on an invention; for example, patent X may cite prior patents A and B, but A may be more important to X’s function than B. Moreover, the reason patent X may cite A and B is to indicate that they are alternative products or methods of production, so that X is not infringing them, rather than that patent X builds on them. Renewals are also observed later in the life of a patent. Another limitation is that renewal decisions may be made not on grounds of patent value, but on changes in the cost of renewing or maintaining patents. The limitation of patent family size as a measure of patent value is that it does not take into account the attractiveness of different market destinations.

Our contribution here is to join recent work that weights patent family size by indicators of the market potential of the countries comprising the patent family (see van Pottelsberghe de la Potterie and van Zeebroeck, 2008; Frietsch et al., 2010; Ernst and Omland, 2011; Kabore, 2012; and Neuhausler and Freitsch, 2013). The rationale for factoring in the market size is that rights holders who possess more valuable patents would self-select in protecting their inventions in greater and larger world markets. Our main value added to this literature is that we develop a novel way to test the extent to which patent family size, weighted by market size, is informative about patent value.

---

8 Also, in Bessen’s (2008) study, citations are found to explain a small percentage of the variation in patent value.

9 van Pottelsberghe de la Potterie and van Zeebroeck (2008) focus on the size and age of a patent family, and discuss how the market size (GDP) of destination countries is an important factor in the patent validation strategy of firms but do not construct a GDP-weighted measure of patent families. Ernst and Omland (2011) weight patent counts by market coverage in order to develop a Patent Asset Index for firms. Specifically, a firm’s patent portfolio is evaluated based on the GDP of countries covered and its technological relevance based on citations received. Freitsch et al. (2010) incorporate export volumes and intensities to patent value indicators. Neuhausler and Freitsch (2013) use six different variables for weighting patent family sizes: the imports, GDP, population, strength of patent protection, global competitiveness, or intensity of local competition of the countries in the family. A key difference between our paper and Neuhausler and Freitsch (2013) is that we weight patent families by the level of GDP so as to capture the absolute market size of a country, whereas Neuhausler and Freitsch (2013) normalize their family size measures (so as to focus on average family sizes); for example, they use the share of a country’s GDP in the world’s total (or for the variables that are indexes, they use the percentage of the maximum value). Kabore (2012) also uses the absolute size of an economy as weights, but conduct different methods of analysis, namely probit and logistic survival analysis.
value. Such a test and demonstration of the predictive ability of market-weighted patent family size has not been done thus far. A key advantage of the weighted patent family size as a measure of patent value is that it is ex ante; that is, it provides more current information about patent value. This would be highly useful for current business decision-making or policy-making, as well as for applied research that utilizes patent data. To demonstrate this advantage, our empirical test centers on how well the weighted patent family size method can forecast patent life and forward citations, which are common measures of patent value after the fact.

The paper is organized as follows. The next section discusses the construction of our measure of patent value and the alternative measures of value. Section III discusses our dataset. Section IV discusses our methodology for testing and comparing our approach to alternative measures of value. Section V contains our results, robustness checks, and an application to show how the method could be used in studies of patent protection. Section VI concludes.

II. VALUING PATENTS BY THE MARKET SIZE OF PATENT FAMILIES

Several reasons exist as to why patent rights are sought in large markets and why such rights are more valuable in those markets. First, the value of obtaining a patent for a new innovation depends on multiple factors, such as the nature of the technology, the inventive step, the demand for that technology by consumers and potential licensees, imitation risk, and the availability of alternative appropriation mechanisms. In small economies where the market for the technology and expected returns are limited, or where imitation risk is low, the incentives for obtaining a patent are relatively small or non-existent. In larger economies, markets for technology tend to be larger (which can both raise the demand for licensing as well as attract piracy and infringement) and competition more tense (owing to rivals that can innovate or imitate and thereby create competing

---

10 The focus here is on patent value and not patent quality. Quality deals more with whether a technology should have been patented; whether it is novel and robust, or correctly awarded. Value deals with the commercial or market value of having a right. Even a minor innovation may be valuable if one can ‘own’ it, particularly if it is overly broad and generic (e.g. algorithm for sorting, where it would indeed be quite profitable to charge other people for the right to use). See Kappos and Graham (2012) for a discussion of measures to improve patent quality.

11 Studies exist which suggest that piracy is a function of market size. Pirates tend to target for duplication products with large mass appeal. For work on copyrighted products, see De Vany and Walls (2007).
technologies). Consequently, in larger economies, the incentives to procure patents and enforce them would tend to be greater.

Second, not only is it costly to file patents in the multiple countries that comprise a patent family, but the cost is higher in larger markets, such as Europe, Japan, and the U.S. Figure 1 shows that the cost of patenting varies positively with the income level of the destination: the larger the market, the greater the cost per patent. Using data for 2010, the figure breaks down destinations by four quartiles, based on GDP (in PPP international dollars), and shows the average fees of obtaining a patent in each income quartile. The fees refer to official fees as well as to agent (or attorney) fees. The fees exclude renewal fees, but an earlier study by Berrier (1996) confirms that the cradle-to-grave costs (i.e., during the lifespan of a patent right) tend to be greater in larger markets, such as Japan, EPO, and U.S. Furthermore, Berger (2005) finds that the cost of patenting in the EPO is compounded by requirements for translation of the patent document and validation fees (whereby an EPO patent must be activated in each individual member state that is designated in the patent). Hence, factoring in the cost and expense of international patent filing, the relatively most valuable innovations would tend to be selected for patenting in the larger markets.

To capture this, we start with a patent family, and weight each country in the family by its market size, namely its gross domestic product. A patent family consists of the priority patent and its subsequent patents. Under the Paris Convention, a priority patent application is an initial patent application that allows an applicant to file subsequent patent applications for the same invention in other countries within a year. Within each patent family, we focus on those countries in which the patent has been granted and validated.

Suppose that $V_j^*$ is the unobserved value of a priority patent $j$. Let $V_j$ be an observable indicator of this value; that is, $V_j^* = V_j^*(V_j)$ is some function of the observable. For example, $V_j$ could be the number of citations received by priority patent $j$ within the first $T$ years of its life.

The data are from Park (2010) and sources cited therein.
Alternatively, it could be the number of countries in the patent family; namely, the country of priority patenting and countries of subsequent patenting. Let the raw family size (count) be:

\[ V_j = \sum_{n=1}^{N} I_n \]  

(1)

where \( I_n \) is an indicator function which equals one if priority patent \( j \) is patented in country \( n \) and zero otherwise, and \( N \) is the total number of countries (including the country of priority filing). Equation (1) is what de Rassenfosse et al. (2014) call ‘geographic’ family size. The more countries covered by the priority patent, the more valuable it may be. However, the raw family size in (1) does not take into account the market sizes of countries. Larger markets are likely to be not only more attractive for exploiting an invention but also more competitive for acquiring patent rights and costlier to enforce them. Hence, we modify (1) to take into account market size:

\[ V_j = \sum_{n=1}^{N} \omega_n I_n \]  

(2)

where \( \omega_n = GDP_n \) (real gross domestic product of country \( n \)). We refer to (2) as the GDP-weighted family size of patent \( j \), which takes into account the composition of the international patent family in terms of the market size of countries.

More specifically, we sample the GDP-weighted family size of priority patent \( j \) at the time of its application (time \( t \)):

\[ V_{jt} = \sum_{n=1}^{N} \omega_{nt} I_{nt} \]  

(3)

That is, we evaluate the market sizes of all the countries in the patent family as of time \( t \). Thus, once we evaluate the global market size of the patent family at time \( t \), we hold that value fixed in our subsequent ex ante analyses of patent value. We do not update the GDPs of the countries in the patent family after time \( t \), as our goal is to evaluate the ex-ante ability of the GDP-weighted patent family size to predict the patent value of the priority patent.
We can then assess this patent valuation ex post by judging how long the patent lived on the premise that patent holders will maintain their more valuable patents. Hence, priority patent \( j \) is renewed in country \( n \) at some future time \( \tau > t \) if the benefit of having a patent right over not having it exceeds the cost of maintaining and enforcing the property right:

\[
R_{n\tau}^j(V^*) - N R_{n\tau}^j(V^*) \geq \kappa_{n\tau}
\]  

and not renewed otherwise, where \( R_{n\tau} \) is the benefit of renewing the patent in country \( n \) at time \( \tau \) and \( NR_{n\tau} \) the benefit of exploiting the technology without the protection of a patent in country \( n \) at time \( \tau \), and \( \kappa_{n\tau} \) denotes the cost of renewing the patent right in country \( n \) at time \( \tau \). In the empirical section, the duration of patent \( j \) is defined as the period between the date of filing and the date when the patent is no longer renewed in any of its family countries.

III. DATA SOURCES

We utilize a dataset of patent families, each of which is composed of the priority patent and subsequent patents that been granted. Data on this come from PatStat 4.21 (the April 2011 version) and the Patent Renewal Status (PRS) database. PatStat is a database gathered by the European Patent Office (EPO) on behalf of the OECD Taskforce on Patent Statistics. PatStat contains files of information on patents from national patent offices and regional patent offices, such as the European Patent Office (EPO). The database covers a long period of time and contains several million observations from around 160 countries in the world. The PRS dataset records events about the life of patents. If a patent has been renewed or has lapsed, PRS will record the event. PatStat does not indicate this particular piece of information. Hence, we merged

---

13See Schankerman and Pakes (1986) and Ernst et al. (2010) for more detailed analyses of the patenting decision. In the empirics, we do not explicitly model the costs of filing and renewing patent rights since we rely on revealed preference; namely, that the costs did not exceed the benefit of obtaining and maintaining rights for those patents still in force.

14For example, suppose a patent is filed in 1980 and has a patent family of three countries. In the first country, the patent is not renewed in 1985, in the second it is not renewed in 1990, and in the third it is not renewed in 1995. The patent’s duration is then 15 years (1995 minus 1980). Van Zeebroeck (2007) calls this the single renewal approach. One advantage of this definition is that it is much less influenced by the maintenance fees of an individual country.
the PRS legal status dataset with the priority patents dataset of PatStat in order to retain patent priorities that have renewal information. Other variables are added, such as forward citations, by matching the merged dataset with the PatStat files that contain the needed variables, as described below.

The patents in this dataset are granted patents. Our empirical analysis covers the cohorts of priority patents associated with patent application years 1980 to 2001 (inclusive). We use the application date instead of, say, the publication date because for patents that have a family size of more than one jurisdiction, the application date of a subsequent filing is the same as the application date of the priority filing, but that is not the case with publication dates. Ending our sample in 2001 allows us to observe the outcomes of various metrics of patent valuation, such as patent citations received in the first few years of patent life beyond 2001.

Our sample also consists of priority patents that have expired by 2009 so that we can eliminate censored observations, namely patents whose ultimate lifespan we do not know about. This results in a sample of priority filings near the end of the sample period (2001) that have relatively lower lifespans. We deal with any potential bias from this later in the paper.

(i) Construction of the Dataset

We extracted over a million patent priority filings from five major source (applicant) countries: USA, Japan, Germany, France, and the United Kingdom. We considered all destination countries in which there could be subsequent filings, including those five countries. Our data include first filings at the EPO and via the Patent Cooperation Treaty (PCT) system. For these filings, we identified the applicant country as the priority country, and to measure their associated market size, we considered the GDP of the member states in which the patent was activated or validated in

---

15Our U.S. priority patent applications date from 1981, given that maintenance fees in the U.S. started in December 1980. Patents in force prior to that were exempt from such fees.
16To clarify, our sample period is 1980-2001, but the PatStat version we use has patent applications data up to 2009.
17The appendix discusses some discrete-time survival analysis that we conducted which helps to treat these censored observations without eliminating them.
the national phase. The data span all eight broad technology classes based on the International Patent Classification (IPC) codes A - H. Figure 2 shows the share of patents across the five source countries. Most of the patents are from the U.S., followed by Germany.

For each of the patent priorities we extracted, we constructed an indicator of patent value based on patent citations, two on patent families, and one on patent duration. Again, the duration of a patent equals the length of time between the date of its application and the date when it lapsed in all of the countries in the patent family. Figure 3 shows the distribution of patents in our sample by their duration. The median duration is between 9 and 10 years of patent life.

The citations-based indicator is the number of citations (including self-citations) received by a priority patent during its first 5 years after publication. If a patent is a single filing, the citations counted are for this singleton. If the patent is a priority filing, the priority receives citations made to it and to any citations made to any subsequent filing. This is also the case when a subsequent filing experiences an opposition: the priority will be considered as having experienced the opposition. As Hall et al. (2001) discuss, the longer the time passes, the more opportunity a patent has to be cited, and so as to minimize this bias we consider citations over this fixed length (first 5 years). In any case, in the dataset, citations tend to drop substantially after 6 - 8 years of publication. Now, the number of citations received by a patent is not readily provided in PatStat. The raw data indicate when a patent cites another patent. However, we have had to derive forward citations (i.e., citations by later patents) by using the publication identification numbers of the citing and cited patents. For example, given a set A of patents that cited patents in set B, it is possible to reverse count how many citations each patent in set B received from

\[ \text{For EPO grants, we summed the GDP of the three leading countries: Germany, France, and U.K., as information about the countries of patent validation was not readily available. These are the three leading countries of patent validation in the EPO (see Figure 2 of Harhoff et al. (2009) (based on an EPO database, EPASYS), and \url{https://www.epo.org/about-us/annual-reports-statistics/annual-report/2009.html} p. 58).} \]

\[ \text{See \url{http://www.wipo.int/classifications/ipc/en/} for more details. In our sample, the top three technologies classes represented are performing operations, human necessities, and physics. Textiles and paper account for the least.} \]

\[ \text{In preliminary analyses, we also worked with citations received during the first 8 years after publication and found the results to be highly similar. The results are available upon request.} \]
patents in set A, using the publication identification number of the cited patents (from set B), along with the publication year data in set B. For instance, to count the citations received during the first 5 years, we used the year of publication of the citing patent as the year the citation was made and attributed it to the year of publication of the cited patent.

The family based indicators include the raw measure, which simply counts the number of countries in the international patent family. This may be as small as one (if, say, only the home country is protected) or as large as 29 (for a few of the priority patents in our sample). The other family size measure is one that sums up the GDP of all the countries in the patent family. GDP is in 2005 constant U.S. dollars. We refer to this measure as the GDP-weighted family size measure. Data on GDP are from the World Bank’s World Development Indicators.

In our empirical analysis, we restrict the sample to priority patent grants that have a family size of at least two countries. Otherwise, in an unrestricted sample, the vast majority of observations represent domestic-only patents; that is, priority patents filed only in one country – typically, the home country – and therefore have a family size of one. An especially large number of these patents are where the priority filing is in the U.S. The U.S. is a very large market – its GDP is typically greater than the sum of that of the other four nations – and so domestic U.S. patents would still be valuable on that basis, even if their family sizes are just one. Thus, a potential problem is that some patents with no subsequent filings may send mixed messages: they could represent patents of modest value or high value if they cover a very large single market. Another matter is that for Japan we only have international patents (where the family size \(>1\)). For these reasons, we focus our analyses on those patent families whose family size exceeds one. We later provide a robustness check where we include the domestic-only patents.

(ii) Summary Statistics

Tables 1 - 2 present some descriptive statistics. The first table shows the means and standard

\[\text{In preliminary analyses, we also worked with private GDP (netting out government expenditures) and found the results also to be very similar. The results are available upon request.}\]
deviations of the indicators broken down by applicant country. The entries in Table 1 show the statistics per patent. The priority patent grant is the unit of analysis. Part A of the table covers patent families of all sizes in the sample, while part B is restricted to those families consisting of at least two countries. It is the latter sample that we focus on in this paper, unless indicated otherwise. Comparing parts A and B, we see that excluding singletons (i.e., patent families consisting of only one country) raises the mean citations received by a priority patent. It also raises the mean market-weighted family size of the patent, except in the case of the U.S. The reason is that the U.S. is the largest market in the sample, so that dropping singletons for the U.S. reduces the average market size of a U.S. patent family. Based on citations, duration, and family size weighted by GDP, Japan appears to have patent priorities with the highest mean patent value. The U.S. has the lowest in terms of duration and Germany the lowest in terms of citations.

For purposes of comparison, Table 1 includes a dummy variable representing opposition. Opposition equals one if the priority patent was challenged or litigated, or had any claims against it by a third party. The patent may be challenged if it is valuable and a rival inventor believes that he already owns the rights to the technology. Rivals may be less inclined to challenge trivial patents or patents that do not have much commercial prospects. Indeed, the table shows that the rate of opposition is greater for international patents. However, this variable may indicate not only whether a patent has much commercial value but whether it is invalid from a legal perspective. For this ambiguity, we do not utilize this variable as an alternative measure of patent value. Table 1 shows another dummy variable that classifies a triadic patent, namely a priority filing that was subsequently filed in the U.S., EPO, and Japan. These patents may be especially highly valuable since they are patented in large markets (see also Kumazawa and Gomis-Porqueras, 2012). But only a small percentage of patents are triadic, and like the raw family size measure, the triadic variable does not take into account the potential market size of the protected areas.

In Table 2, we break the restricted sample into four groups (quartiles) based on duration: patents at the bottom 25% of duration (or length of life), the next 25%, the second top 25%, and
the top 25%. The tables show that during the time the patent is in force, the duration of a patent and the citations it receives are positively correlated. This raises some issues about the endogeneity between them, which we address later on. For family size weighted by GDP, we observe that longer lived patents tend to have had greater GDP in their patent families at the time of application. The table also shows that opposition is not a consistent measure of duration. Patents that are in the second quartile have lower opposition rates (35%) than those in the bottom quartile (46%), but patents in the top quartile have the highest opposition rates (55%).

Figure 4 provides a preview of the tests and analyses in the next section. We sorted patents in ascending order of value according to the GDP-weighted family size method and then formed quartiles. We did the same using the citations method of counting the number of citations in the first five years. Then, for each quartile, we asked what percentage of patents in that group survived for 12 years or longer. Using either method of valuation, we find that more than 25% of patents in each of the top two quartiles live relatively long lives. However, the citation method can underestimate lifespan. Almost 41% of patents deemed to be in the bottom quartile – receiving relatively the least citations – do end up having long patent lives.

IV. METHODOLOGY: FORECASTING EXERCISE

Our task is to compare among the various indicators of patent value and conduct a kind of ‘horse race’ to see which one might contain the better signal of patent value. First, we compare the GDP-weighted patent family size measure against patent citations obtained within the first 5 years, to predict the duration of patents. Second, we compare the GDP-weighted patent family size measure against the duration of patents to predict patent citations received within the first five years.

To carry out our test of predictive properties, we need an estimation period and an out-of-sample forecasting period. To obtain these, we randomly divided our sample of priority patents into two balanced groups. Specifically, we used the stratified random sampling method on six
major strata: the countries of origin, the IPC code, patent duration, the application filing year, oppositions, and whether the patent is a triadic one. The sampling is done across all six strata to ensure that the two resulting subsets are representative of the overall population.\footnote{We need to perform this stratification because otherwise any random cut of the data might put nearly all of one country’s data in one sample, or put nearly all of one field of technology in another, which will lead to biased results.}

Let us call one sample $z = 0$ and the other sample $z = 1$. We then have an equation for patent duration that contains one of the patent value indicators as one of the independent variables. The methodology here is to estimate this model using the $z = 0$ dataset, and then to use the estimated model (or fitted model) to predict the duration in the $z = 1$ sample, and finally to compare the actual duration of the $z = 1$ sample against the forecasted duration of the $z = 1$ sample. The model, with whichever indicator of patent value, that has the best forecast accuracy (for example, the lowest root mean square error) is deemed to be the best at predicting duration. For robustness, we perform the reverse: we estimate the model using the $z = 1$ dataset and use it to predict duration in the $z = 0$ dataset. This methodology is intended to mimic how the GDP-weighted patent family size measure might be used in reality; namely, to predict at time $t$ (the present), how long the patent will live (an ex-ante perspective). The reason we do not perform this exercise (i.e., to forecast duration and then compare to actual duration) with the original (whole) sample is that we would be using the same patent data to make the predictions that we used to estimate the parameters. By splitting the sample into two sub-samples ($z = 0$ and $z = 1$), we use an estimated model from one dataset (one realization of the world) to forecast duration in another draw of the data. In other words, we create separate environments in which to estimate and test the model, thereby creating conditions similar to out-of-sample forecasting.

To recap, we estimate the following equations on the $z = 0$ sub-sample:

\[ T = \alpha + \beta C + \gamma X + \varepsilon \]  \hspace{1cm} (5)

\[ T = \alpha + \beta F + \gamma X + \varepsilon \]  \hspace{1cm} (6)
\[ T = \alpha + \beta V + \gamma X + \varepsilon \]  

Equation (5) uses \( C \), the number of citations by the 5th year, along with a vector of control variables, \( X \), which include fixed effects and interaction effects, to predict duration \( T \). Equation (6) uses \( F \), the raw family size measure (simple counts of countries in the family), instead of \( C \), to help predict \( T \), and equation (7) uses \( V \), the GDP-weighted family size measure, to help predict \( T \).

We then obtain the following “fitted” equations:

\[
\hat{T} = \hat{\alpha} + \hat{\beta} C + \hat{\gamma} X \tag{8}
\]

\[
\hat{T} = \hat{\alpha} + \hat{\beta} F + \hat{\gamma} X \tag{9}
\]

\[
\hat{T} = \hat{\alpha} + \hat{\beta} V + \hat{\gamma} X \tag{10}
\]

and apply them to dataset \( z = 1 \). That is, using the estimated \( \alpha \), \( \beta \), and \( \gamma \) from the \( z = 0 \) sample, we plug in the data for \( C \), \( F \), \( V \), and \( X \) from the \( z = 1 \) sample to generate predicted values \( \hat{T} \) and compare them to the actual values of duration \( T \) in the \( z = 1 \) sample.

As measures of forecast accuracy, we use two kinds. First, the Root Mean Squared Percentage Error:

\[
RMSPE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left( \frac{\hat{T}_i - T_i}{T_i} \right)^2}
\]

where \( N \) is the sample size of the \( z = 1 \) sample. The \( i^{th} \) subscript refers to the \( i^{th} \) patent in the \( z = 1 \) sample. This gives us the average forecast errors \( (\hat{T}_i - T_i) \) as a percentage of the actual value.

Another measure of forecast accuracy is Theil’s inequality coefficient (the U1 version):

\[
U = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^{N} (\hat{T}_i - T_i)^2}}{\sqrt{\frac{1}{N} \sum_{i=1}^{N} (\hat{T}_i)^2} + \sqrt{\frac{1}{N} \sum_{i=1}^{N} (T_i)^2}}
\]

where \( U = 0 \) if there is a perfect fit. Again, for robustness, we will repeat the entire exercise.
by reversing sub-samples: using the \( z = 1 \) sample for estimation and \( z = 0 \) for prediction. For further tests of robustness, we will change the dependent variable of interest from duration to citations, and determine how well the ex-ante GDP-weighted family size measure of a priority patent predicts the citations that a patent will receive in five years’ time and compare that to the ability of a patent’s duration to predict the number of citations it receives in the first five years.

V. RESULTS

Our first ‘horse race’ test involves predicting the duration of patents and the second involves predicting the citations received. To control for the many unobservable factors, our estimation equations include a full set of source (applicant) country fixed effects, year fixed effects, and technological field fixed effects (given by their IPC codes). Specifically, we have five country dummies (France, Germany, Japan, United Kingdom, and United States); twenty-two year dummies (for annual data from 1980 to 2001); and eight IPC dummies (corresponding to single letter codes A to H). We also control for country and year interaction effects to control for factors that vary by country and year; technological field and year interaction effects to control for factors that vary by IPC and year; and technological field and country interaction effects to capture factors that vary by country and IPC. These fixed effects and interaction effects should capture differences in technologies, policy regimes, patenting costs, and research productivity, among others.

(i) Forecasting

Table 3 reports on the fitted equations for predicting the duration of patents. The models estimated using the \( z = 0 \) sub-sample are in columns 1 - 3 and those estimated using the \( z = 1 \) sub-sample are in columns 4 - 6. Recall that we use the equations shown in columns 1 - 3 to perform out-of-sample forecasting on the \( z = 1 \) dataset and use those in columns 4 - 6 to do the same on the \( z = 0 \) dataset. We also estimated a model (not shown) that is similar to those in Table 3 except that it does not include any of the patent value indicators. We refer to this as the ‘base case’. If it were the case that the base case models predict the best, this would cast doubt
on the usefulness of all our indicators of patent value. As the table shows, the coefficient estimates of the patent indicators are all significant at the 1% level.

Table 4 contains the results of the ‘horse race’ tests. It summarizes the ability of the various patent value indicators at forecasting patent duration, using the two measures of forecast accuracy discussed earlier: the root mean squared percentage error (RMSPE) and Theil’s U. Panel (i) reports on the use of the estimated model from the \( z = 0 \) sample to predict duration in the \( z = 1 \) sample, and panel (ii) shows the reverse. For each of these panels, the two different forecast accuracy values are reported, and these values are also normalized with respect to the forecast accuracy value associated with the GDP-weighted family size measure (which is shown in column 3). This helps us better see how the other patent value indicators perform relative to our proposed GDP-weighted family size measure.

Based on both forecast accuracy criteria, the model which uses the GDP-weighted family size variable predicts the best, followed by the citations variable. Essentially, using the RMSPE criterion, GDP-weighted family size is 6.6% more accurate at predicting duration than citations received the first 5 years. Using the Theil criterion, we find that GDP weighted family size predicts patent duration more accurately than citations by 4.8%. If we do not weight patent families by GDP, family size is not as accurate at predicting duration as are citations. Indeed, it gives marginally better forecasts than the base case scenario of not using any indicator of patent value. While the advantage of raw patent family size is its simplicity – easy to observe and apply – it ignores the market sizes of the countries in the patent family and does not have strong predictive power. Throughout, the worst performance (at predicting how long a patent lives) is given by models that do not use any of these patent value indicators.

Of course, the forecasting models could all be improved by combining the various patent indicators, such as citations and GDP-weighted patent family size, and entering them all in the forecasting equations. But our objective was to conduct a ‘horse race’ test and so it was easier to show the usefulness by entering these patent indicators one at a time. The more important reason,
however, is that in practice, citations are not observed ex ante. At the time of a patent priority application, we do not observe citations until some years later. We do, nevertheless, observe family size and the GDP levels of countries in the patent family. Hence, a chief advantage of the GDP-weighted patent family size measure is that it utilizes information that is contemporaneously available with patent filings.

The horse race exercise we conducted focused on patent duration as the benchmark for patent value. We now turn to patent citations as the benchmark of value, and see how the GDP-weighted family size measure performs against duration at predicting citations received by a patent. The significance is to show that the GDP-weighted family size measure has broader influences on patent value beyond patent renewal behavior, since duration is not a perfect measure of value either (e.g., duration can be longer if the costs of renewing patents are lower, and there are no changes in the benefits or value of patents).

Table 5 shows the estimates of the equations used to predict the future citations of patents, 5 years ahead after first filing the patent. We use family size as well as the weighted family size variables, along with fixed effects and interaction effects between country and time, technological field and time, and country and technological field as the independent variables. We also use the duration of a patent as a predictor of the number of citations it receives. Even though a patent can be cited long after the right expires or lapses, duration and citations may be correlated during the lifetime of a patent, since the more time passes, the more citations the patent may receive. We will deal with this simultaneity issue shortly, after we first see how family size fares relative to the duration variable as a predictor of citations, assuming that duration has the inherent advantage of being correlated with citations. We should expect each of the predictors to perform better than a model where none of them are present, namely the base case.

What we find is that the coefficients of these predictors are all statistically significant. Each can contribute to predicting citations. As before, to determine which can perform relatively better, we estimate the model using one split sample to forecast citations in the other, and then vice versa.
In contrast to the models for predicting duration, the models here do not fit as well: the adjusted R-squared’s are considerably lower. There seems to be a lot of noise in patent citations that is hard to capture. That said, Table 6 summarizes the measures of forecast accuracy. Again, the GDP-weighted family size variable performs best at predicting future citations. What is quite interesting is that it performs even better than the duration of patents at predicting citations.

ii) Robustness Checks

As mentioned earlier, a potential issue with the results in Tables 5 and 6 is that of endogeneity between duration and citations during the time the patent is in force; that is, more citations are likely to have been received in a patent’s fifth year of life than after its first year (and the same patent may keep on receiving citations long after it expires). But note that our citation counts are based on a fixed duration. We have effectively stopped the citations ‘clock’ at the fifth year. Thus, to avoid possible simultaneity between citations and duration, we can focus on patents that lived longer than five years. We illustrate with the case of patents that lived longer than five years and the citations they received during the first five years. This way, we can use the eventual duration of the patent to see if it signals high patent value in terms of the number of cites it receives in the first five years.

We then re-estimated the type of models shown in Table 5. To conserve space, we do not report the regression results (which are available upon request), but focus on discussing the measures of forecast accuracy. Scenario A of Table 7 contains the results (which can be compared to Table 6). The results show again that the GDP weighted family size measure performs better than duration at predicting citations received.

Another bias to address is the potential effect of ending the sample period in 2001. We have had to drop those patents that did not expire by 2009, the last year for which we had data available, since we do not know their eventual duration. This may cause our sample of priority filings near the end of the sample period (2001) to be biased towards low to medium quality since
some of them will have lived for 9 years or less. Some of this heterogeneity by cohort is controlled for by the year effects and their interactions with other variables. But to make the dataset more balanced throughout the sample period, we can restrict the dataset to patents that lived no more than 9 years. Table 7 shows that in the case of predicting durations (see Scenario Bi), the results are qualitatively the same as before: GDP-weighted family size has the highest forecast accuracy. Similarly in the case of predicting citations (see Scenario Bii), GDP-weighted family size performs best, even against patent duration.

Another robustness check is to include domestic-only patents in our sample; that is, priority patents filed only in the home country and which, as a result, have a family size of one. Scenario C of Table 7 summarizes the forecast accuracies under this unrestricted sample. To conserve space, we again do not show the estimation equations but show only the normalized forecast accuracy measures – that is, the RMSPE (or Theil U) of each indicator relative to the RMSPE (or Theil U) of the GDP-weighted family size indicator. Even for this full sample of patent families of all sizes, the GDP-weighted family measure predicts duration the best among the alternative patent value indicators in either sub-sample (z = 0 or z = 1).

(iii) Application to Equivalent Subsidy Rate (ESR)

Thus far, we have focused on constructing and comparing alternative measures of patent value. The next step would be to apply these measures and relate them to say, measures of productivity or R&D, among other variables. A thorough treatment of the potential applications requires a longer analysis and is beyond the scope of this paper. Instead, we provide a preview of a potential application and suggest provisionally how our measure of patent value relates to aggregate business enterprise R&D (BERD) performed. Following Schankerman (1998), Bessen (2008), and Arora et al. (2008), we derive the equivalent subsidy rate (ESR) of patent protection. As Schankerman (1998) first explained, an equivalent subsidy is one that would incentivize firms, in the absence of

\[23\] A more thorough treatment could estimate, for example, the short run effects, long run dynamics, and lag structures underlying the patent-R&D relationship.
patent protection, to do the same R&D that they would do with patent protection. The ESR is derived as the ratio of the value of patents to R&D. The intuition is that the numerator represents the value of having a patent right compared to having no such right (recall the LHS of equation 4). The ESR is then this value as a percentage of R&D. This ESR should be viewed as a kind of ‘upper bound’ in that it assumes that inventions have value only if they are patented.

We derive the mean ESR at the national level. Ideally, we should be matching each priority patent to the patent owner’s R&D. As Bessen (2008) noted, aggregate measures of ESR do not accurately reflect the subsidy that a firm would need to conduct R&D. Our goal here is merely illustrative. To derive the country-wide measure of patent value using the GDP-weighted family size method, we calculate the following: \( V_t^i = \frac{\sum_{j=1}^{J(i,t)} V_{jt}}{\sum_{j=1}^{J(i,t)} (\sum_{n=1}^{N} GDP_{nt} I_{nt})} \) for \( i = \) France, Germany, Japan, UK, and USA. \( J(i,t) \) is the count of priority patents in country \( i \) at time \( t \). Here \( V_t^i \) measures the value of the \( i^{th} \) country’s flow of priority patents at time \( t \), where each patent is evaluated as the sum of the market sizes in which it has patent protection. The indicator variable \( I_{nt} \) = 1 if the priority filing \( j \) includes destination \( n \) in the patent family at time \( t \), and zero otherwise. We can interpret the ratio \( V_t^i \) to R&D as the market coverage per $1 of R&D. It is the required access to markets that yields the same R&D effort in the absence of patent protection.

Table 8 presents some sample equivalent subsidy rates by applicant country and by select technological fields. Column 2 shows the mean value of a country’s patents, where each patent is weighted by the GDP of its international patent family. In total, U.S. patents have the greatest global market coverage, followed by Japan and then Germany. The average U.S. patent, however, has the smallest market coverage among these countries, consistent with Table 1, part B. The average Japanese patent has the greatest coverage (see Column 3). Column 4 shows the level of industrial R&D by country and column 5 the equivalent subsidy rate (ESR) calculations, namely

\[ V_t^i = \frac{\sum_{j=1}^{J(i,t)} V_{jt}}{\sum_{j=1}^{J(i,t)} (\sum_{n=1}^{N} GDP_{nt} I_{nt})} \]

\[ I_{nt} = 1 \text{ if the priority filing } j \text{ includes destination } n \text{ in the patent family at time } t, \text{ and zero otherwise.} \]

\[ \text{To clarify, } N \text{ refers to all the countries in the world, beyond the five source countries. The priority patents are all granted. Time } t \text{ is their date of patent application. In accordance with what we have done thus far, we measure their market size at the time of application.} \]

\[ \text{As earlier, we restrict the sample to international patents, where family size } > 1. \]
the mean value of patents (based on the GDP-weighted family size method) per R&D. The calculations indicate that the ESR is relatively highest in Germany; that is, patent protection is relatively most important there and relatively least important in the U.S. Further study is needed to explain these trends. Among the factors might be that alternative means of protecting innovation are more easily available to U.S. firms, like reputation or technological complexity, or that U.S. companies have first-mover advantages and commanding sales after entry into regional and global markets. Columns 6 - 8 break the ESR calculations down by technological field: Chemistry, Physics, and Electricity. For these calculations, patent values were divided by industry-specific R&D. The concordance is not perfect, but we used approximate measures as a first start. The calculations indicate the ESR to be greatest among Electricity patents, followed by Physics, and then by Chemistry. This is consistent with Schankerman (1998) who also finds that patent protection is more important for electronic and mechanical patents than for chemical or pharmaceutical ones. Again, more research is needed, but differences in industrial regulation may help explain the differences in ESR among technologies, as well as differences in the source of returns to innovation.

VI. CONCLUSION

In this paper, we constructed an indicator of patent value based on a patent’s family size and composition, weighting each country in the family by its market size (namely GDP). This approach overcomes the disadvantage of existing measures of patent valuation which yield information about patent value after the fact; for example, after a patent’s record of citations has been observed over a period of time, or after a patent right has been observed to lapse or be renewed at a later point in time. To the extent that our GDP-weighted patent family measure is a good ex ante indicator of patent value, then ex post, patents with greater GDP-weighted patent family values should live longer and have greater citations (over some specified period of time). This was the motivation for conducting the forecasting exercises; namely, to see whether our measure of interest

---

26 R&D data are obtained from the OECD’s Main Science and Technology Indicators.

27 For example, R&D in chemicals (including pharmaceuticals) were matched to patenting in Chemistry, R&D in computers, optics, and electronic products to patenting in Physics, and R&D in electrical equipment and machinery to patenting in Electricity.

21
can predict the valuable patents that we would observe ex post. To do this, we had split the sample into two groups, estimated a forecasting model over one group, and then used the other group, treating it as another draw of data, to conduct out-of-sample forecasts. We then reversed the situation, estimating the model over the latter group and conducting forecasting exercises with the first group. Our main result is that our measure of patent valuation out-performs the citations method at predicting patent duration and renewal, and out-performs patent duration at predicting citations. And it improves upon models that do not use any indicator of patent value.

Of course, this is not to say that ex post valuation matters less. In many cases, especially for conducting welfare analyses, it is useful to know whether patents granted in earlier periods turned out to be valuable and to provide feedback on the policies that gave rise to them (or failed to). But where anticipating ‘value’ is crucial for making current resource allocation decisions, the valuation method discussed here serves to help minimize forecast errors.

As extensions to this paper, the GDP-weighted patent family size measure can be used to create indicators of patent values by industry and/or country, and they in turn can be applied to studies on productivity, R&D, or technology diffusion. However, in estimating the total value of national patents $V^i_t$, it will be useful to control for changes in the propensity to patent, which may not remain stable over time. The GDP-weighted family size measure may potentially fluctuate because of shifts in the patenting strategies of firms. An option is to focus on the average value of national patents over time or between countries (i.e., $\frac{V^i_t}{J_{i,t}}$, where $J$ denotes the count of national priority patents). Future work could also go one step further and weight the market sizes of countries by some measure of commercial or technological development. In other words, two or more countries with equal GDPs may still differ with respect to other dimensions that pertain to valuation.
APPENDIX: DISCRETE TIME SURVIVAL ANALYSIS

As an alternative approach, we could have focused on when a patent expires, rather than its length of patent life. Survival analysis has the advantage of dealing with censored observations; for example, patents in the sample that have not yet expired or whose lapse date we do not observe.\textsuperscript{28} We find that this alternative methodological approach also supports our method of patent valuation. As the table below shows, when we control for fixed effects and interaction effects, the citations variables have the ‘wrong’ signs in that they predict that the greater the citations, the greater the likelihood of the lapse of patents. In contrast, the GDP-weighted patent family variable has the expected negative coefficient.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
<td>(6)</td>
</tr>
<tr>
<td>(\ln(\text{Family Size}))</td>
<td>-0.261***</td>
<td>0.052***</td>
<td>0.052***</td>
<td>-0.092***</td>
<td>-0.003***</td>
<td>-0.003***</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.049)</td>
<td>(0.049)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>(\ln(\text{Weighted Fam Size}))</td>
<td>-0.092***</td>
<td>0.006***</td>
<td>0.006***</td>
<td>0.006***</td>
<td>0.006***</td>
<td>0.006***</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>(\ln(1 + \text{Citations}))</td>
<td>-0.151***</td>
<td>-0.003***</td>
<td>-0.003***</td>
<td>-0.003***</td>
<td>-0.003***</td>
<td>-0.003***</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>Country Fixed Effects</td>
<td>Excluded</td>
<td>Excluded</td>
<td>Excluded</td>
<td>Included</td>
<td>Included</td>
<td>Included</td>
</tr>
<tr>
<td>Year Fixed Effects</td>
<td>Excluded</td>
<td>Excluded</td>
<td>Excluded</td>
<td>Included</td>
<td>Included</td>
<td>Included</td>
</tr>
<tr>
<td>Technology Fixed Effects</td>
<td>Excluded</td>
<td>Excluded</td>
<td>Excluded</td>
<td>Included</td>
<td>Included</td>
<td>Included</td>
</tr>
<tr>
<td>LR Chi (2)</td>
<td>4.13E+06</td>
<td>4.12E+06</td>
<td>4.13E+06</td>
<td>4.30E+06</td>
<td>4.30E+06</td>
<td>4.30E+06</td>
</tr>
</tbody>
</table>

Observations = 10,616,465. Given a coefficient of the hazard rate, \(\alpha\), the corresponding odds ratio \(\beta = e^\alpha\). A negative coefficient estimate indicates a decrease in the hazard of lapse. Standard errors in parentheses.

\(***\) p<0.01, ** p<0.05, * p<0.1

On a technical note, the above estimation assumed that the date of expiry is normally distributed (which can be questioned) and that the process for survival is not memory-less. The conditional probability of “lapse” in a given short interval was allowed to depend on the number of years the patent has been in force, so that the hazard function varied. We also made no assumptions about whether the hazard function exhibits positive or negative duration dependence. The analysis also required us to expand the number of observations greatly into a format suitable for discrete time survival analysis.\textsuperscript{29}

\textsuperscript{28}See Marco (2007), Zeebroeck (2007), and Kabore (2012) for related survival analyses of patents and innovation.

\textsuperscript{29}More details about our patent survival analysis are available upon request.
References


Table 1: Sample Statistics of Patent Value Indicators, by Source Country

### A. All Patent Families

<table>
<thead>
<tr>
<th>Source Country</th>
<th>Duration (Years)</th>
<th>Family Size</th>
<th>Weighted Family Size</th>
<th>Citations first 5 yrs</th>
<th>Opposition (yes, no)</th>
<th>Triadic Patent</th>
</tr>
</thead>
<tbody>
<tr>
<td>France</td>
<td>8.85</td>
<td>1.42</td>
<td>4.52e+12</td>
<td>0.72</td>
<td>0.76</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>(3.99)</td>
<td>(0.53)</td>
<td>(3.87e+12)</td>
<td>(1.60)</td>
<td>(0.43)</td>
<td>(0.13)</td>
</tr>
<tr>
<td>Germany</td>
<td>8.78</td>
<td>1.28</td>
<td>3.86e+12</td>
<td>0.54</td>
<td>0.54</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(3.66)</td>
<td>(0.48)</td>
<td>(3.26e+12)</td>
<td>(1.37)</td>
<td>(0.50)</td>
<td>(0.09)</td>
</tr>
<tr>
<td>Japan</td>
<td>9.97</td>
<td>1.30</td>
<td>7.43e+12</td>
<td>1.24</td>
<td>0.49</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>(3.22)</td>
<td>(0.61)</td>
<td>(3.84e+12)</td>
<td>(2.20)</td>
<td>(0.50)</td>
<td>(0.19)</td>
</tr>
<tr>
<td>U.K.</td>
<td>8.85</td>
<td>1.48</td>
<td>5.77e+12</td>
<td>0.90</td>
<td>0.69</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>(3.68)</td>
<td>(0.59)</td>
<td>(3.73e+12)</td>
<td>(2.04)</td>
<td>(0.46)</td>
<td>(0.16)</td>
</tr>
<tr>
<td>U.S.A.</td>
<td>8.29</td>
<td>1.21</td>
<td>6.59e+12</td>
<td>0.76</td>
<td>0.29</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(2.76)</td>
<td>(0.48)</td>
<td>(2.78e+12)</td>
<td>(1.39)</td>
<td>(0.45)</td>
<td>(0.09)</td>
</tr>
<tr>
<td>Total</td>
<td>8.79</td>
<td>1.27</td>
<td>5.89e+12</td>
<td>0.80</td>
<td>0.44</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(3.28)</td>
<td>(0.52)</td>
<td>(3.51e+12)</td>
<td>(1.63)</td>
<td>(0.50)</td>
<td>(0.12)</td>
</tr>
</tbody>
</table>

### B. International Patent Families: Family Size > 1

<table>
<thead>
<tr>
<th>Source Country</th>
<th>Duration (Years)</th>
<th>Family Size</th>
<th>Weighted Family Size</th>
<th>Citations first 5 yrs</th>
<th>Opposition (yes, no)</th>
<th>Triadic Patent</th>
</tr>
</thead>
<tbody>
<tr>
<td>France</td>
<td>9.96</td>
<td>2.04</td>
<td>7.11e+12</td>
<td>1.11</td>
<td>0.89</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>(4.10)</td>
<td>(0.23)</td>
<td>(3.78e+12)</td>
<td>(2.12)</td>
<td>(0.32)</td>
<td>(0.19)</td>
</tr>
<tr>
<td>Germany</td>
<td>9.79</td>
<td>2.04</td>
<td>6.68e+12</td>
<td>1.10</td>
<td>0.82</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>(4.14)</td>
<td>(0.24)</td>
<td>(3.22e+12)</td>
<td>(2.09)</td>
<td>(0.38)</td>
<td>(0.17)</td>
</tr>
<tr>
<td>Japan</td>
<td>11.07</td>
<td>2.17</td>
<td>1.05e+13</td>
<td>1.90</td>
<td>0.66</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>(3.60)</td>
<td>(0.68)</td>
<td>(4.41e+12)</td>
<td>(3.13)</td>
<td>(0.48)</td>
<td>(0.33)</td>
</tr>
<tr>
<td>U.K.</td>
<td>9.52</td>
<td>2.08</td>
<td>7.07e+12</td>
<td>1.28</td>
<td>0.83</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>(3.97)</td>
<td>(0.39)</td>
<td>(3.68e+12)</td>
<td>(2.72)</td>
<td>(0.38)</td>
<td>(0.22)</td>
</tr>
<tr>
<td>U.S.A.</td>
<td>8.21</td>
<td>2.10</td>
<td>6.28e+12</td>
<td>0.92</td>
<td>0.83</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>(3.33)</td>
<td>(0.47)</td>
<td>(3.52e+12)</td>
<td>(2.49)</td>
<td>(0.38)</td>
<td>(0.19)</td>
</tr>
<tr>
<td>Total</td>
<td>8.79</td>
<td>1.27</td>
<td>5.89e+12</td>
<td>0.80</td>
<td>0.44</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(3.90)</td>
<td>(0.45)</td>
<td>(3.99e+12)</td>
<td>(2.54)</td>
<td>(0.40)</td>
<td>(0.22)</td>
</tr>
</tbody>
</table>

Notes: Unit of analysis is a patent. The table shows means, with standard deviations in parentheses. Number of observations in panel A = 1,208,345 and in panel B = 299,540. Sample period is 1980-2001. Duration of patent life is measured from the date of application to the date of its last non-renewal. Family size is the count of countries associated with a priority patent. Weighted family size is the count weighted by the country’s GDP in real 2005 U.S. dollars. Citations are counts of forward citations by other patents during the first 5 years of a patent’s life. Opposition is a dummy variable = 1 if a patent was opposed by a third party after it was granted. Triadic Patent = 1 if a patent is protected simultaneously in the U.S., Japan, and European Patent Office (EPO).
Table 2: Patent Valuation Indicators Grouped by Duration Rank

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>27.4</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>2</td>
<td>23.6</td>
<td>1.655</td>
<td>1.010</td>
<td>1.077</td>
<td>1.636</td>
</tr>
<tr>
<td>3</td>
<td>28.4</td>
<td>2.281</td>
<td>1.034</td>
<td>1.553</td>
<td>3.432</td>
</tr>
<tr>
<td>4</td>
<td>20.6</td>
<td>3.138</td>
<td>1.079</td>
<td>1.726</td>
<td>5.386</td>
</tr>
</tbody>
</table>

The table shows the means normalized by the mean of the first quartile. Number of Observations = 299,540 and time period is 1980 - 2001. See Table 1 for variable definitions.

Table 3: Regression Model for Predicting Duration, by Split Samples

<table>
<thead>
<tr>
<th>Split Sample</th>
<th>ln(Family Size)</th>
<th>ln(Weighted Fam Size)</th>
<th>ln(1 + Citations 1st 5 yrs)</th>
<th>Observations</th>
<th>Adj R-squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>z=0</td>
<td>0.461***</td>
<td>0.318***</td>
<td>0.136***</td>
<td>166,799</td>
<td>0.970</td>
</tr>
<tr>
<td>z=1</td>
<td>0.413***</td>
<td>0.270***</td>
<td>0.117***</td>
<td>132,735</td>
<td>0.975</td>
</tr>
</tbody>
</table>

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1
Sample is restricted to family size > 1. Included in each regression are: Country Fixed Effects, Year Fixed Effects, Technology Fixed Effects, Country x Year Interactions, Technology x Year Interactions, Country x Technology Interactions. Coefficient estimates of all the fixed effects and interaction terms are suppressed to conserve space.
Table 4: Measures of Forecast Accuracy

**i. Using the z=0 sample to predict duration in the z=1 sample**

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Additional RHS Variables:</td>
<td>Base Case</td>
<td>(No Value Indicators)</td>
<td>Family Size</td>
<td>Weighted Family Size</td>
</tr>
<tr>
<td>RMSPE</td>
<td>0.234</td>
<td>0.231</td>
<td>0.213</td>
<td>0.227</td>
</tr>
<tr>
<td>- Ratio to Col. (3)</td>
<td>1.099</td>
<td>1.085</td>
<td>1.000</td>
<td>1.066</td>
</tr>
<tr>
<td>THEIL U</td>
<td>0.081</td>
<td>0.080</td>
<td>0.076</td>
<td>0.079</td>
</tr>
<tr>
<td>- Ratio to Col. (3)</td>
<td>1.061</td>
<td>1.048</td>
<td>1.000</td>
<td>1.048</td>
</tr>
</tbody>
</table>

**ii. Using the z=1 sample to predict duration in the z=0 sample**

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Additional RHS Variables:</td>
<td>Base Case</td>
<td>(No Value Indicators)</td>
<td>Family Size</td>
<td>Weighted Family Size</td>
</tr>
<tr>
<td>RMSPE</td>
<td>0.267</td>
<td>0.264</td>
<td>0.245</td>
<td>0.260</td>
</tr>
<tr>
<td>- Ratio to Col. (3)</td>
<td>1.094</td>
<td>1.082</td>
<td>1.000</td>
<td>1.065</td>
</tr>
<tr>
<td>THEIL U</td>
<td>0.088</td>
<td>0.087</td>
<td>0.081</td>
<td>0.086</td>
</tr>
<tr>
<td>- Ratio to Col. (3)</td>
<td>1.085</td>
<td>1.073</td>
<td>1.000</td>
<td>1.061</td>
</tr>
</tbody>
</table>

Notes: RMPSE denotes Root Mean Squared Percentage Error and THEIL U is Theil’s measure of forecast accuracy. Case i is based on estimated models 1-3 in Table 3 and Case ii on models 4-6 in Table 3. Column 1 shows the forecast accuracy associated with not using any of the patent value indicators. Columns 2-4 show the accuracies associated with adding one of the indicators shown.
Table 5: **Regression Model for Predicting Citations Received**
*First 5 Years, by Split Samples*

<table>
<thead>
<tr>
<th>Split Sample</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$z=0$</td>
<td>$z=0$</td>
<td>$z=0$</td>
<td>$z=1$</td>
<td>$z=1$</td>
<td>$z=1$</td>
</tr>
<tr>
<td>ln($Family\ Size$)</td>
<td>0.697***</td>
<td>0.582***</td>
<td>0.513***</td>
<td>0.488***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.016)</td>
<td>(0.018)</td>
<td>(0.004)</td>
<td>(0.004)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln($Weighted\ Fam\ Size$)</td>
<td></td>
<td></td>
<td>0.352***</td>
<td></td>
<td>0.338***</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.004)</td>
<td></td>
<td>(0.005)</td>
<td></td>
</tr>
<tr>
<td>ln($Duration$)</td>
<td></td>
<td></td>
<td></td>
<td>0.498</td>
<td>0.476</td>
<td>0.536</td>
</tr>
<tr>
<td>Observations</td>
<td>166,805</td>
<td>166,805</td>
<td>166,799</td>
<td>132,735</td>
<td>132,735</td>
<td>132,735</td>
</tr>
<tr>
<td>Adj R-squared</td>
<td>0.484</td>
<td>0.549</td>
<td>0.498</td>
<td>0.476</td>
<td>0.536</td>
<td>0.489</td>
</tr>
</tbody>
</table>

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1
Sample is restricted to family size > 1. Included in each regression are: Country Fixed Effects, Year Fixed Effects, Technology Fixed Effects, Country x Year Interactions, Technology x Year Interactions, Country x Technology Interactions. Coefficient estimates of all the fixed effects and interaction terms are suppressed to conserve space.
Table 6: Measures of Forecast Accuracy for Predicting Citations Received in the First 5 Years

### i. Using the $z=0$ sample to predict citations in the $z=1$ sample

<table>
<thead>
<tr>
<th>Additional RHS Variables:</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base Case (No Value Indicators)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSPE</td>
<td>0.514</td>
<td>0.519</td>
<td>0.498</td>
<td>0.513</td>
</tr>
<tr>
<td>– Ratio to Col. (3)</td>
<td>1.031</td>
<td>1.040</td>
<td>1.000</td>
<td>1.028</td>
</tr>
<tr>
<td>THEIL U</td>
<td>0.450</td>
<td>0.448</td>
<td>0.432</td>
<td>0.445</td>
</tr>
<tr>
<td>– Ratio to Col. (3)</td>
<td>1.042</td>
<td>1.037</td>
<td>1.000</td>
<td>1.030</td>
</tr>
</tbody>
</table>

### ii. Using the $z=1$ sample to predict duration in the $z=0$ sample

<table>
<thead>
<tr>
<th>Additional RHS Variables:</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base Case (No Value Indicators)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSPE</td>
<td>0.517</td>
<td>0.518</td>
<td>0.502</td>
<td>0.516</td>
</tr>
<tr>
<td>– Ratio to Col. (3)</td>
<td>1.030</td>
<td>1.033</td>
<td>1.000</td>
<td>1.028</td>
</tr>
<tr>
<td>THEIL U</td>
<td>0.453</td>
<td>0.450</td>
<td>0.433</td>
<td>0.447</td>
</tr>
<tr>
<td>– Ratio to Col. (3)</td>
<td>1.046</td>
<td>1.038</td>
<td>1.000</td>
<td>1.032</td>
</tr>
</tbody>
</table>

Notes: RMSE denotes Root Mean Squared Percentage Error and THEIL U is Theil’s measure of forecast accuracy. Case i is based on estimated models 1-3 in Table 5 and Case ii on models 4-6 in Table 5. Column 1 shows the forecast accuracy associated with not using any of the patent value indicators. Columns 2-4 show the accuracies associated with adding one of the indicators shown.
Table 7: Robustness Checks

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Forecast Accuracy:</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
<th>(9)</th>
<th>(10)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Base Case Family Size Weighted Family Size Duration 1st 5 years</td>
<td>RMSPE</td>
<td>THEIL U</td>
<td>RMSPE</td>
<td>THEIL U</td>
<td>RMSPE</td>
<td>THEIL U</td>
<td>RMSPE</td>
<td>THEIL U</td>
<td>RMSPE</td>
<td>THEIL U</td>
</tr>
<tr>
<td>A.</td>
<td>RMSPE</td>
<td>1.027</td>
<td>1.038</td>
<td>1.000</td>
<td>1.033</td>
<td></td>
<td></td>
<td>1.031</td>
<td>1.034</td>
<td>1.000</td>
<td>1.029</td>
</tr>
<tr>
<td></td>
<td>THEIL U</td>
<td>1.041</td>
<td>1.034</td>
<td>1.000</td>
<td>1.032</td>
<td></td>
<td></td>
<td>1.047</td>
<td>1.036</td>
<td>1.000</td>
<td>1.034</td>
</tr>
<tr>
<td>Bi.</td>
<td>RMSPE</td>
<td>1.025</td>
<td>1.023</td>
<td>1.000</td>
<td></td>
<td>1.019</td>
<td></td>
<td>1.024</td>
<td>1.022</td>
<td>1.000</td>
<td>1.018</td>
</tr>
<tr>
<td></td>
<td>THEIL U</td>
<td>1.010</td>
<td>1.009</td>
<td>1.000</td>
<td></td>
<td>1.006</td>
<td></td>
<td>1.016</td>
<td>1.014</td>
<td>1.000</td>
<td>1.013</td>
</tr>
<tr>
<td>Bii.</td>
<td>RMSPE</td>
<td>1.027</td>
<td>1.026</td>
<td>1.000</td>
<td>1.018</td>
<td></td>
<td></td>
<td>1.023</td>
<td>1.021</td>
<td>1.000</td>
<td>1.025</td>
</tr>
<tr>
<td></td>
<td>THEIL U</td>
<td>1.081</td>
<td>1.080</td>
<td>1.000</td>
<td>1.071</td>
<td></td>
<td></td>
<td>1.081</td>
<td>1.079</td>
<td>1.000</td>
<td>1.076</td>
</tr>
<tr>
<td>C.</td>
<td>RMSPE</td>
<td>1.042</td>
<td>1.043</td>
<td>1.000</td>
<td></td>
<td>1.033</td>
<td></td>
<td>1.049</td>
<td>1.049</td>
<td>1.000</td>
<td>1.039</td>
</tr>
<tr>
<td></td>
<td>THEIL U</td>
<td>1.025</td>
<td>1.021</td>
<td>1.000</td>
<td></td>
<td>1.018</td>
<td></td>
<td>1.034</td>
<td>1.028</td>
<td>1.000</td>
<td>1.024</td>
</tr>
</tbody>
</table>

Entries show measures of forecast accuracy normalized by the accuracy measure of the GDP-weighted family size indicator.

Scenarios:

A. Addresses potential endogeneity between duration and citations. Predicts citations (1st 5 yrs) using sample of patents whose duration > 5 yrs.

Bi. Restricts sample to patents whose duration ≤ 9 yrs. Forecast variable is duration.

Bii. Restricts sample to patents whose duration ≤ 9 yrs. Forecast variable is citations (received 1st 5 yrs).

C. Includes patents of family size = 1 in the sample. Forecast variable is duration.
Table 8: **Sample Equivalent Subsidy Rates, by Source Country and Technology Field**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>France</td>
<td>7360</td>
<td>7.26</td>
<td>20585</td>
<td>383</td>
<td>66</td>
<td>386</td>
<td>140</td>
<td></td>
</tr>
<tr>
<td>Germany</td>
<td>18700</td>
<td>6.74</td>
<td>37348</td>
<td>549</td>
<td>567</td>
<td>287</td>
<td>595</td>
<td></td>
</tr>
<tr>
<td>Japan</td>
<td>21600</td>
<td>8.97</td>
<td>70068</td>
<td>332</td>
<td>458</td>
<td>307</td>
<td>157</td>
<td></td>
</tr>
<tr>
<td>U.K.</td>
<td>6220</td>
<td>6.92</td>
<td>18485</td>
<td>356</td>
<td>138</td>
<td>901</td>
<td>335</td>
<td></td>
</tr>
<tr>
<td>U.S.A.</td>
<td>23300</td>
<td>6.35</td>
<td>176290</td>
<td>145</td>
<td>245</td>
<td>899</td>
<td>692</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>15300</td>
<td>7.24</td>
<td>64555</td>
<td>355</td>
<td>293</td>
<td>398</td>
<td>152</td>
<td></td>
</tr>
</tbody>
</table>

Notes: The sample period is 1980-2009. Value of national patents equals the GDP of the countries covered by all of a country’s patents over this period. The sample of patents is restricted to those whose family size > 1. Industry R&D is business enterprise research and development performed. ESR denotes equivalent subsidy rate and is the mean ratio of the value of patents (column 2) to R&D (column 4). Chemistry, physics, and Electricity refer to IPC codes C, G, and H respectively. ESR is interpreted as billions of market coverage per dollar of R&D.
Figure 1. Relationship between Patenting Cost and Market Size

Figure 2. Percentage Distribution of Patents by Country

Official and Associate Fees in 2010, by income group (GDP, International PPP dollars)