ECARES

Productivity and Propensity:

The Two Faces of the R&D-Patent Relationship

Gaétan de Rassenfosse
SBS-EM, ECARES, Université Libre de Bruxelles and CEPR

ECARES working paper 2010-025

ECARES
ULB - CP 114
50, F.D. Roosevelt Ave., B-1050 Brussels BELGIUM
www.ecares.org

Productivity and propensity: The two faces of the R&D-patent relationship

Gaétan de Rassenfosse*

June 1, 2010

Abstract

This paper tackles one of the most persistent criticism of patent statistics. Because not all inventions are patented, the patent-to-R&D ratio reflects both a productivity effect (the number of inventions created per unit of research input) and a propensity effect (the proportion of inventions patented). We propose a solution to this identification problem. Our methodology uses information on the density of patent value and leads to results that are easy to interpret. It is applied to a novel data set of priority patent applications in which each patent is fractionally allocated to its inventors' countries and to the technological areas to which it belongs. Interestingly, it is frequently observed that an industry may exhibit a low number of patents per unit of R&D in one country yet actually be more productive than the same industry in another country where the patent-to-R&D ratio is higher.

Keywords: identification strategy, patent family, patent value, research productivity, propensity to patent
JEL Classification: O30, O52

*Université libre de Bruxelles (Solvay Brussels School of Economics and Management, ECARES), FNRS Research Fellow. The author is also member of ECORE, the association between CORE and ECARES. Email: gderasse@ulb.ac.be. The author is indebted to Bruno Cassiman, Michele Cincera, Jérôme Danguy, Catherine Dehon, Keld Laursen, Carine Peeters, Bruno van Pottelsberghe, Nicolas van Zeebroeck and Andrea Vezzulli for helpful comments and discussions. The paper has also benefited from comments of participants of various seminars including the Summer School in Economics of IPR (Bologna), the KUL PhD Workshop (Toulouse) and the 2nd Patent Data Workshop at IPTS (Sevilla).

1 Introduction

This paper tackles one of the most persistent criticisms of patent statistics. The major drawback of patent indicators is certainly that they are heavily influenced by varying patent practices across companies, industries and countries such that patent indicators do not perfectly signal innovation performances. Not all inventions are patentable, and not all patentable inventions are patented. As a result, the patent-to-R&D ratio reflects both a *productivity* effect (which leads from research efforts to inventions) and a *propensity* effect (which leads from inventions to patents). Information on the productivity of research efforts conveyed by patents is noisy and comparison across units should be undertaken with care. In the words of Lanjouw et al. (1998, p. 413), "*one of the longest lasting debates in the history of economic measurement has been whether the noise and the biases in patent count measures can be made small enough to make patent counts useful measures of innovative output in economic studies.*"

Separating the productivity effect from the propensity nuisance is subject to much empirical complexity and no convincing method has been proposed so far. Two remedies are usually applied to filter out the noise in patent statistics. A first practice consists in selecting only valuable patents, such as the patents filed at a regional or an international office (e.g., the European Patent Office (EPO) or the United States Patent and Trademark Office (USPTO)), or the patents filed simultaneously at several offices. An important indicator in this respect is the count of triadic patents put forward by the OECD (Dernis et al., 2001). Triadic patents are patents filed simultaneously at the Japan Patent Office (JPO), the USPTO and the EPO. Since these patents are international by nature, national patent practices have only a limited influence on the count. A second remedy consists in weighting patent data with some value indicators, such as the patent family size or the number of forward citations the patent received (e.g., Schankerman and Pakes, 1986). Yet, all existing techniques have their limitations and one can expect better statistics in light of the available data.

This paper puts forward a methodology to explicitly decompose the patent-to-R&D ratio into its components of productivity and propensity. In so doing, it also sheds new light on existing patent indicators. For instance, the count of triadic patents can be seen as a special case of the model. The methodology is designed for cross-country analyses of patenting performances at the industry level. Under two general assumptions, we show that the density of patent value can be used to identify the two effects. The assumptions are that (i) the density of invention value is universal and (ii) the probability of patenting inventions of high enough value is similar across countries. An easy-to-implement criterion to select the appropriate value threshold is then derived. The methodology leads to results that have a clear interpretation and should be appealing to both scholars and policymakers. In particular, it can be used to remove the noise in patent statistics due to varying propensities to patent across countries.

An empirical application of the model is proposed. It is applied to a novel data set of priority patent applications for six industries in four countries (Belgium, Germany, France and the Netherlands) over the period 1988 to 1992. Priority filings are collected regardless of the patent office of applications and are fractionally allocated to their inventors' countries and to the technological areas to which they belong. A key ingredient in the identification procedure is patent value, which is measured with three indicators of patent family size. Interestingly, the results suggest that more patents per R&D are not necessarily reflective of a higher productivity of research. Indeed, it is frequently observed that an industry may exhibit a low number of patents per R&D in one country yet actually be more productive than the same industry in another country where the patent-to-R&D ratio is higher.

The paper is organized as follows. The next section reviews the literature on innovation indicators in general, and patent indicators in particular. Section 3 presents the model and discusses the assumptions. An empirical application is presented in Section 4 and the results are analyzed in Section 5. Finally, Section 6 concludes.

2 Patent statistics as economic indicators

Patent is only one of many innovation indicators, but it is probably the one that has attracted most of the attention by scholars.[1] One can say that patent data together with scientific publication data belong to a first group of indicators that capture the *outputs of the innovation process*. A great feature of patent and publication data is that they contain much information that is semi-structured such that it is possible to carry out detailed investigations of the innovation process. For instance, citations to a patent document or a publication can be used to study knowledge spillovers or assess depreciation rates of knowledge (Jaffe and Trajtenberg, 1996; McDowell, 1982). This first group also encompasses data collected at trade fairs or in trade journals, which are particularly useful in studying innovation in small companies (see, e.g., Coombs et al., 1996, for an attempt to create an innovation indicator based on product announcements in trade journals). A second group of innovation indicators includes the *inputs to the innovation process* such as R&D expenditures or the number of researchers. This data can typically be broken down according to the source of funding (private or public) and to the use (basic research, applied research or development work). However, it is silent as to how efficiently the research inputs are used to produce innovation. It also captures only a fraction of total expenditures devoted to innovation. Using detailed data for the Netherlands in 1992, Brouwer and Kleinknecht (1997) have estimated that R&D expenditures represent about one quarter of total innovation expenditures. The largest share of the "unmeasured" expenditures consists of investment in fixed assets. A third group of innovation indicators include all the *observable effects of innovation activities* such as a country's export share of high technology prod-

[1]A comprehensive discussion on the measurement of scientific output as well as an historic perspective is provided in Godin (2005). Patel and Pavitt (1995) have a good description of the major measures of technological activity.

ucts or a firm's share of sales that is due to new products. A popular source of information in this respect is the European Community Innovation Survey (CIS) which asks information on the share of sales with products that are new to the firm (as a measure of imitation) or new to the market (as a measure of innovation). Kleinknecht et al. (2002) have shown that R&D expenditures were highly correlated with the number of patents, but that both of these measures were much less correlated with the level of innovative sales. In other words, the latter measure conveys an additional information that is not captured with traditional innovation indicators (see, e.g., Cassiman and Veugelers, 2002; Crépon et al., 1998; Czarnitzki and Toole, 2010 for studies that use the CIS data). These data are very useful to collect information on process innovations or on innovation in services, which are more difficult to capture with traditional innovation indicators.

The richness of patent data as well as their wide availability explain the success of patent indicators. Besides, the information is available for long time series and for a large number of countries. Yet, statistics based on patent data are not free of pitfalls and biases. We shall briefly discuss four of them.[2]

First, many aspects need to be considered when building a patent indicator. These include the geographical allocation of the invention (country of residence of inventor vs. country of residence of applicant), the legal status of the document (applications vs. granted patents, priority applications vs. second filings), and the period of reference (application date vs. grant date). The geographical scope of the indicator also matters. Are patents counted at one reference patent office (such as the EPO or the USPTO), at each national patent office (such as the French patent office for French applicants and the German patent office for German applicants), or is a worldwide scope desired? Similarly, should the indicator be a straight count of individual patents or a more subtle measure based on patent family such as triadic or foreign-oriented patents? (See Dernis et al., 2001; WIPO, 2008.) All these technical choices make patent indicators difficult to understand for the uninitiated.

Second, patents vary greatly in their technological and economic significance. While some patents protect a general purpose technology, some others are merely minor improvement of existing technologies. Besides, some patents are clearly harebrained (think of the spork, a combined spoon and fork, or a walking through walls training system) or worse, should never have been granted. For instance, many observers believe that Amazon's controversial "one-click" patent should never have been granted by the USPTO (Jaffe and Lerner, 2004). In a similar vein, the same patent application may have a different outcome in different jurisdictions (Jensen et al., 2006; Mejer and van Pottelsberghe, 2009). It is no surprise that the distribution of patent value is constantly found to be highly skewed to the left, with a majority of patents having no or low value (see, e.g., Grabowski and Vernon, 1994; Pakes, 1986). Evidence by Scherer and Harhoff

[2]See Griliches (1990) and Schmookler (1954) for complementary discussions on the use and limits of patent statistics as economic indicators. See Dernis et al. (2001), Grupp and Schmoch (1999), OECD (2009), Pavitt (1985), Soete and Wyatt (1983) for interesting thoughts on the use of patent statistics to assess countries' innovation performances.

(2000) from eight sets of data on invention attributable to private sector firms and universities suggests that the top 10% most valuable observations capture from 48% to 93% of total sample returns.

Third, patent indicators miss various non-patented inventions, which probably account for (much) more than 50% of the inventions. Investment in research activity leads to inventions that may, in turn, lead to patent applications. In a study of patent practices of U.S. manufacturing firms, Cohen et al. (2000) report mean patenting rates of around 50% for product innovations in the early nineties. Arundel and Kabla (1998) obtain rates lower than 40% for Europe's largest industrial firms at around the same period. Using data from an international survey performed in 2005, de Rassenfosse (2010) finds that propensity rates average around 50%.

Fourth, a somewhat related issue is the identification problem in patent statistics: They do not allow distinguishing between the *productivity effect* that leads from R&D to inventions and the *propensity effect* that leads from inventions to patents. Patent-to-R&D ratio is indeed determined by both the number of inventions per unit of R&D and the proportion of these inventions that are patented. This concern has been raised by many scholars (see, e.g., Griliches, 1990; Kortum, 1993; Lanjouw and Schankerman, 2004) but has been only partially addressed in the literature. Yet, the propensity to patent has been found to affect patent statistics to a significant extent. de Rassenfosse and van Pottelsberghe (2009) take into account determinants of both the productivity and the propensity components in modeling the R&D-patent relationship. They find evidence that patent statistics reflect both effects, though to varying extent according to the patent indicator that is used. Whereas priority filings are mostly determined by propensity variables, triadic patents are more reflective of the productivity of research.

Scholars have proposed different techniques to get rid of the noise that affects patent statistics. A first, commonly used, practice consists in selecting only the patents whose value exceeds a given threshold. This is done by counting patents filed at one single reference office such as the EPO or the USPTO — it is presumably more expensive for, say, a Belgian applicant to file at the USPTO or at the EPO in lieu of the Belgian patent office — or by selecting patents which have a broad geographical coverage such as triadic patents. Soete and Wyatt (1983) and Grupp and Schmoch (1999) are early examples of this approach. A second way towards meaningful patent statistics has been proposed by Schankerman and Pakes (1986). It consists in correcting straight counts by use of value indicators. Schankerman and Pakes (1986) use patent renewal data to estimate the value of patent rights and find that much of the decline in patents per scientist and engineer over the period 1965 to 1975 disappears when patent counts are adjusted for quality. Other value indicators were also tried successfully. Trajtenberg (1990) presents patent counts weighted by total citations while Tong and Frame (1994) use the number of claims instead of patents to measure innovativeness. Lanjouw et al. (1998) and van Pottelsberghe and van Zeebroeck (2008) propose a simple weighting scheme based on data on family size and renewal. Lanjouw and Schankerman (2004) and van Zeebroeck

(2010) go one step further: they use multiple indicators to improve the observed value of patents.

However useful they may be, these techniques have their limits. For instance, the count of EPO or USPTO patents favors European or North-American countries (the so-called home bias). The presence of a systematic bias has been detected even for European applicants filing patents at the EPO. de Rassenfosse and van Pottelsberghe (2007) have shown that applicants in countries that are early signatories to the European Patent Convention are more likely to transfer their patents to the EPO, suggesting that a learning effect in using the European patent office exists. Triadic patents are less subject to geographical or institutional biases. However, the apparent burst in innovation suggested by the increase in the number of triadic patents is explained by a greater propensity to "go triadic" (Danguy et al., 2009). In other words, triadic patents do not consistently measure productivity effects over time. The practice that consists in using value indicators to weight patents is also subject to criticism. Most notably, it is difficult to choose a convincing weighting scheme. Any weights are arbitrary even though they are found to be optimal by some measure. Tong and Frame (1994) propose weights that are linear in the number of claims. Trajtenberg (1990), for his part, finds that there are increasing returns to information content of citations and that the weights should increase more than proportionally with citations.

3 The Model

We present a methodology to identify the productivity and the propensity effects in the R&D-patent relationship. The input (the research efforts) and the patented output are observed but the object of interest (inventions) is not. As a result, it is not possible to disentangle the productivity effect, which leads from R&D to inventions, from the propensity effect, which leads from inventions to patents. The identification relies on a model of the patent generating process where inventions created are patented with a probability that depends on their market value. The methodology requires a large enough amount of data and works if patent practices are similar to some extent. It is therefore best suited for cross-country analyses at the industry-level, as opposed to firm-level analyses where the number of observations may not be sufficient or patent practices may be too different. In other words, the methodology can be used to compare all the patents pertaining to a given industry in one country vis-à-vis all the patents pertaining to the same industry in another country.

The number of inventions (in a given industry) of country i is a linear function of the research effort R_i and the productivity of research λ_i, the rate at which inventions are discovered. It is thus given by $R_i \lambda_i$, which is unobservable. We allow heterogeneity in the private value of inventions. We note the density function of the probability distribution of invention value by $f_i(v)$. The probability of patenting an invention of value v is denoted by $p_i(v)$. The overall propensity to patent δ_i is thus given by

$$\delta_i = \int_0^\infty f_i(x) p_i(x) dx \tag{1}$$

The total number of patents applied, P_i, is given by

$$P_i = \delta_i \lambda_i R_i \tag{2}$$

Obviously, there exists an infinite number of combinations (δ, λ) for any given P/R ratio. Since δ and λ are unknown, equation (2) explicitly characterizes the identification problem. The objective is to identify the true (δ, λ) couple. Unfortunately, as long as the functional form of $f(v)$ and the associated probability to patent are unknown, absolute values for the propensity and the productivity components cannot be recovered. In order to disentangle the productivity from the propensity component, two identifying assumptions need be introduced. The first, denoted A1, is that *the distribution of invention value is universal.* That is, a Dutch inventor has the same probability than, say, a Belgian inventor to discover a high (or a low) value invention. The absolute number of such high value inventions, however, may differ across countries due to the total amount of resources devoted to research activities (R) and their average productivity (λ). The second assumption, A2, is that *there is a threshold value above which the patenting behavior is similar across units.* In other words, A2 states that the probability of patenting a high value invention is similar across units. The validity of these assumptions is discussed at length at the end of the section. Note however that the identification procedure is designed for cross-country analyses at the industry level, where we believe the assumptions are more likely to hold.

These two assumptions are sufficient to recover relative productivity and propensity rates. Under assumption A1, $f_i(v) = f(v) \; \forall \, i$. Hence, equation (1) can be written as

$$\delta_i = \int_0^\infty f(x) p_i(x) dx = \int_0^\infty d_i(x) dx \tag{3}$$

where $d_i(x)$ can be seen as the distribution of patent value relative to that of invention value (i.e., the distribution of invention value weighted by the probability that patenting occurs). Note that the assumption implies that the distribution of *patent* value across countries is explained by varying probabilities to patent. Note also that if one is able to observe each patent's value, one can actually build the distribution of patent value $D_i(v)$, a normalization of mass 1 of $d_i(v)$. The relative propensity rate between country i and country j is given by

$$r_{i,j} = \frac{\int_0^\infty d_i(x)dx}{\int_0^\infty d_j(x)dx} = \frac{\lambda_i R_i \int_0^\infty d_i(x)dx}{\lambda_j R_j \int_0^\infty d_j(x)dx} \frac{\lambda_j R_j}{\lambda_i R_i} = \frac{P_i}{P_j} \frac{\lambda_j R_j}{\lambda_i R_i} \tag{4}$$

Since the productivity parameters are unknown, equation (4) cannot be estimated directly. Under assumption A2, however, it is possible to computed relative productivity levels. A2 states that $p_i(v) = p(v) \; \forall \; v \geq v_t$, where v_t will be referred to as the *threshold value.* The number of patents with value greater or equal to v_t, $\dot{P}_{i,v_t}$, is given by

$$\dot{P}_{i,v_t} = \lambda_i R_i \int_{v_t}^{\infty} f(x)p(x)dx. \tag{5}$$

Hence, the ratio of productivity parameters can be expressed as

$$\frac{\lambda_i}{\lambda_j} = \frac{\dot{P}_{i,v_t}}{R_i} \frac{R_j}{\dot{P}_{j,v_t}} \tag{6}$$

that is, the ratio of the relative numbers of high value patents. Interestingly, intuition has led scholars to naturally estimate equation (6). For instance, the practice which consists in comparing triadic (or EPO) patents per R&D to evaluate countries' innovation performances can be seen as a special case of the model where the value threshold is set such that only triadic (or EPO) patents are counted in $\dot{P}$. The present model allows understanding the conditions under which such a practice makes sense.

Plugging equation (6) into equation (4) gives a simple expression for the ratio of propensities $r_{i,j}$

$$r_{i,j} = \frac{P_i \dot{P}_{j,v_t}}{P_j \dot{P}_{i,v_t}} \tag{7}$$

Two remarks need to be made regarding equation (7). First, it does not depend on the size of the research efforts; it can be recovered only from the output of the invention process. It follows that it is not impacted by misspecifications of the invention generation process. For instance, if the patent production function of equation (2) exhibits non-constant returns to research efforts, the propensity parameter will not be changed. (Note, however, that the productivity parameter will change.) Second, it is straightforward to compute since the probability to patent needs not be known. The technical challenges lies in identifying v_t correctly in order to compute $\dot{P}$.

Two decision rules can be thought of to determine v_t. The first consists in considering, say, the top 10 or 20 percent most valuable patents in each country to be of high enough value. It is trivial to show, however, that this approach is uninformative since it imposes a constant propensity across countries and leaves all the variation be explained by productivity effects. Indeed, if the top α percent most valuable patents in each country are considered, then the ratio of propensities in equation (7) can be written as $r_{i,j} = (P_i \alpha P_j)/(P_j \alpha P_i) = 1$ and the ratio of productivities in equation (6) as $(\alpha^{-1} P_i R_j)/(R_i \alpha^{-1} P_j) = (P_i/R_i)/(P_j/R_j)$. The second rule would be to select the top 10 or 20 percent most valuable patent across all patent offices (instead of the top at each country), which is equivalent to setting a unique value threshold v_t across countries. In the following, we derive a criterion to determine v_t in a way that is consistent with the model.

Assumption A2 implies that equation (6) not only holds for the interval $[v_t, \infty)$ but also for the interval $[v_t, v_t + \Delta v]$, where Δv is henceforth referred to as the *value step* (or simply *step*). We define

$$\ddot{P}_{i,v_t,\Delta v} = \lambda_i R_i \int_{v_t}^{v_t+\Delta v} f(x)p(x)dx \tag{8}$$

It follows that another expression for the ratio of productivities can be computed using $\ddot{P}_i$ (to simplify notation, $\dot{P}_{i,v_t}$ and $\ddot{P}_{i,v_t,\Delta v}$ are henceforth referred to as $\dot{P}_i$ and $\ddot{P}_i$)

$$\frac{\lambda_i}{\lambda_j} = \frac{\ddot{P}_i}{R_i}\frac{R_j}{\ddot{P}_j} \tag{9}$$

Since the ratios (6) and (9) need to be equal, the value v_t must be such that

$$\frac{\dot{P}_i}{R_i}\frac{R_j}{\dot{P}_j} = \frac{\ddot{P}_i}{R_i}\frac{R_j}{\ddot{P}_j} \tag{10}$$

Or equivalently

$$\frac{\dot{P}_i \ddot{P}_j}{\dot{P}_j \ddot{P}_i} = 1. \tag{11}$$

We will refer to equation (11) as the *identification criterion*. Note that criterion (11) must be true for any $v_t' > v_t$ and for any step Δv. However, choosing a too high threshold value reduces the likelihood to observe any patent at all given the limited number of such patents — especially in small industries where it is common to observe no high value patents in certain years. Hence, the statistical significance of the observed number of patents is likely to be very low. A contrario, there is no penalty in choosing the smallest possible value for v_t such that equation (11) holds. For this reason, equation (11) should be evaluated at $v_t' = v_t$. The question of the optimal Δv is more complex. Choosing a too small step increases the risk of having too few observations to reach statistical significance. On the other hand, a too large step renders the criterion meaningless. By definition, indeed, $\lim_{\Delta v \to \infty} \ddot{P} = \dot{P}$ such that equation (11) always holds for sufficiently large values of Δv. A good practice would thus be to estimate equation (11) with a small step to avoid the statistical artifact created by large Δv's. Under assumptions A1 and A2, equation (11) holds if the probability of patenting is similar. It is important to point out that the reverse is not true. It is not because equation (11) holds that the assumptions are necessarily verified. Nevertheless, it is an encouraging sign that should bring some confidence in the identification procedure.

As previously mentioned, the identification rests on two assumptions. First, the density function of the value of inventions is universal. Second, the probability of patenting inventions of large enough value is similar across countries. There is little evidence in the literature to validate or invalidate these assumptions.[3] However, if the exercise is performed for countries that are at a similar

[3] Needless to say, it would be erroneous to use the distribution of patent value as a proxy for the distribution of invention value to test assumption A1. Such an approach is facially inconsistent with the model.

level of development, there are few reasons to suspect a dramatically different distribution of invention value. Regarding the second assumption, international treaties such as the Paris Convention for the protection of industrial property of 1883 or the Patent Cooperation Treaty of 1978 have set some degree of minimum legal and quality standards across patent systems. In addition, it is more likely that there is a low heterogeneity in the propensity to patent a high value invention than a low value one. A firm that patent few, indeed, will probably select only patents that have a high potential. In contrast, firms that have a high propensity to patent will protect invention of lower marginal value. It is all the more true if the exercise is performed at the industry-level rather than at the country or the firm-level. Finally, it is worth emphasizing that assumption A2 is less restrictive than what is commonly assumed in the literature. Indeed, many models consider that patenting occurs with certainty when the benefits associated with patents exceed the cost of patenting. Such a strong assumption necessarily implies that A2 is fulfilled. In contrast, all is needed for the model to hold is a similar propensity to patent (which may not necessarily be of 1) above a high enough threshold value (which may be well above patenting costs).

The next two sections present an empirical application of the identification procedure. Section 4 explains the patent count and the value indicators that are used, and Section 5 presents the results.

4 Empirical Implementation

The empirical exercise requires detailed information on an industry's patent applications. For this reason, great care is devoted to the construction of the patent indicator. Three methodological choices are discussed. First, the methodology adopted to allocate patents to countries and industries is presented. Second, three measures of patent value are introduced. They are all based on the family size. Finally, the composition of the sample is explained. The analysis of the data thus created is presented in the next section.

Raw data on patents come from the EPO-OECD Patstat database (April 2009 version). Data on R&D expenditures come from the OECD STAN database.

4.1 Patent count

Contrary to most studies that use patent data, we make a distinction between priority patent applications (the first patent protecting an invention) and second filings (which are patent applications filed to extend the initial protection to other jurisdictions). The former serves as the basis for the "raw" count and the latter is used to compute the value of each priority filing. Two allocation criteria can be used to assign patents to countries: the country of residence of inventor(s), or the country of residence of applicant(s). In this study, the paternity of a patent is given on a fractional basis according to the country of inventor(s) as an indicator of the place where the invention was created. This ensures a best match between the measure of input, provided on a territorial basis, and the output. Thus, we survey all priority filings created by inventors of a country, regardless of the patent office of application.[4] This particular counting methodology has been introduced by de Rassenfosse et al. (2010).

An important problem that must be overcome to allocate patents is the missing information on the country of inventors. We use a technique proposed by de Rassenfosse et al. (2010) which has been shown to perform very well. It consists in retrieving data on the country of inventors from different sources of information following a well-defined sequence. First, the missing information is searched for in the potential second filings. If it is found, it is taken from the earliest second filing that contains the information. If no information is available in the second filings, or if there are no second filings, the country of residence of inventors is assumed to be that of applicants. Finally, if no applicant was identified (either in the patent document itself or in its second filings), then the country of inventor is assumed to be that of the country of applications. Once all the information has been retrieved, assignment to countries is performed on a fractional basis. For instance, if a patent has three German inventors and 1 French inventor, the count for Germany will be of 0.75 and that of France of 0.25.

Table 1 shows the office of applications for patents by Belgian, German, French and Dutch inventors over the period 1988 to 1992. Patents are mostly filed in the home office, especially in Germany and France where the national patent office receives more than 90% of all the filings made by inventors in the country. In contrast, patents by Belgian and Dutch inventors are more geographically dispersed. The national patent office of Belgium, for instance, receives less than half of the applications by Belgian inventors and the EPO and the USPTO capture 17% and 9% respectively. The last two columns provide information on missing data on inventors. The figures in the next-to-last column (column A) are the share of patents that initially had missing information on inventors. Information was missing for more than 9 out of 10 patents attributed to French inventors. The last column reports the share of priority filings that were attributed by default to the country of the patent office when it was not possible to recover the information from other sources. Only 6% of the French patents were attributed by default. These patents were filed at the French patent office, had missing information on inventors and applicants, and it was not possible to recover information on inventors from potential second filings.

The allocation to industries is less obvious. Patent offices assign one or more IPC codes to the patents they receive (IPC is the acronym for International Patent Classification). These codes, which are attributed according to the different areas of technology to which the patents pertain, can be used to allocate patents to industries. Two problems must be tackled. First, as for the origin of inventors, IPC codes are frequently missing. We developed an algorithm similar to that for inventors by which missing information is retrieved from second filings. If no information is available in the second filings, or if there

[4] In practice, however, we restrict the data-gathering exercise to 44 patent offices: the patent offices in OECD countries, in EU27 and the BRICS.

Table 1: Origin of priority filings (allocation by the country of inventors).

	Patent office							(A)	(B)
Inv.	BE	DE	FR	NL	EP	US	RoW		
BE	2,513	205	230	148	861	463	706	0.66	0.18
DE	21	129,683	155	41	4,260	2,354	3,722	0.26	0.07
FR	35	426	53,219	19	786	1,048	2,122	0.95	0.06
NL	98	312	98	8,395	1,540	688	1,339	0.89	0.27

Notes: "Inv." and "RoW" stands for "inventors" and "rest of the world" respectively. Column (A) indicates the share of patents that had initially missing information on inventors. Column (B) reports the share of priority filings that were allocated by default to the country of the patent office because it was not possible to identify the country of residence of inventor. Source: PATSTAT April 2009. Own computation. See main text and de Rassenfosse et al. (2010) for methodological details.

is no second filing, the field is left empty. Second, it is necessary that IPC codes be classified according to the International Standard Industrial Classification (ISIC) to ensure correspondence with measures of R&D expenditures. We do so by means of the concordance table between IPC and ISIC codes provided by Schmoch et al. (2003). The authors have estimated an empirical concordance by investigating the patenting activity by technology-based fields (IPC) of more than 3,000 firms classified by industrial sector (ISIC). Similarly to the inventors, the industry allocation is performed on a fractional basis when a patent contains more than one IPC code.[5] Note that other concordance tables exist such as the MERIT concordance table (Verspagen et al., 1994) or the Yale Technology Concordance table (Kortum and Putnam, 1997). We use Schmoch et al. (2003)'s table because it is based on European data and is the most recent table available.

4.2 Value indicators

Patents vary greatly in their economic and technological significance. While some patents are minor improvements to an existing technology, others have significant commercial potential. Several variables have been found to capture patent value or to better reflect innovation performance. For instance, Tong and Frame (1994) provide evidence that patent claims data may be a better indicator of technological effort than straight patent counts. In a similar vein, Lerner (1994) finds that biotechnology patents with larger scope (as measured with the number of IPC classes to which the patent has been assigned) are more valuable to the company that holds them. The citations a patent received is also strongly correlated with its value as shown by Trajtenberg (1990). The very

[5]There are thus two reasons why patents may not be allocated to industrial sectors. First, the information on IPC could not be retrived. Second, the information is available, but the IPC code is not listed in the concordance table. The former case is by far the most frequent.

existence of the citing patents attests to the fact than the cited patent opened the way to a technologically successful line of innovation. Other, more direct, value indicators may also be considered. Lanjouw et al. (1998) use information on the age at which a patent was allowed to lapse and the set of countries in which it was filed to assess its value. In contrast to citations or claims data, the geographical scope of coverage and the number of years the patent was kept in force are associated with monetary costs that reflect applicants' valuation of the patent. In a study of the correlation between the private value of patent rights as estimated by patent holders and various indicators of value, Harhoff et al. (2003) confirm that forward citations do capture patent value but find that patents which are upheld in opposition and annulment procedures and patents representing large international patent families are particularly valuable. In addition to being easily available, information on patent family is thus also a good measure of value. For these reasons, we use family size to measure patent value.[6] Three value indicators are computed.

- ☐ The first indicator measures an application family size as the number of second filings that claim its priority. This indicator, denoted SF, is thus a simple count of all the second filings that derive from the priority filing. It is easy to compute but is only a rough estimate of family size given that one second filing may claim more than one priority document. This indicator is thus subject to potential double counting, which is not accounted for. For instance, if patent P3 claims both the priority patent applications P1 and P2, and P4 claims only P2, the count for P1 will be of two (P1 and P3 are in the same family) and the count for P2 will be of three (P2, P3 and P4 are in the same family). See Appendix A for a schematic representation and further explanation.

- ☐ The second indicator (henceforth $WIDE$) considers a more global measure of family size that corrects for double counting. Patents are said to belong to the same family if they claim at least one priority in common. This family is known as the INPADOC family (INPADOC stands for international patent documentation center). In the above example, the 4 patents belong to the same family, which is of size four. Since there are two priority filings in the family, each priority document is given a weight of 0.5 and a "value" of 4. This measure, however, is subject to another bias that plagues patent families, namely the varying geographical coverage. There is presumably a difference in the value of a family composed of one priority filing and 10 second filings in 10 different countries, as opposed to a family of 10 priority filings in the same country that give rise to only 1 foreign extension. The former is probably more valuable.

- ☐ In order to better reflect the geographical coverage of the family, the third

[6] It must be noted, however, that the international expansion implies that the patent is more valuable but the opposite is not true. Valuable patents are not necessarily extended to foreign markets. This being said, patents that are not extended abroad de facto lose some value as the monopoly power conferred by the patent right is reduced.

indicator measures the family size as the number of countries covered by the family ($GEOG$). For instance, if P1, P3 and P4 are each filed in a different patent office but P2 is filed in the same office as P1, the size of the family will be three.

The distribution of value was also estimated from the citations received by patents. However, research has shown that citation practices differ greatly across patent offices such that international comparisons based on (mostly) national patent citations may induce a strong bias (Harhoff et al., 2008; Meyer, 2000). In addition, missing information induces a pernicious bias: it is difficult to differentiate a patent that has no citation from one that has missing information on citations. As a result, patents with missing information are necessarily associated with low value patents. This issue is of particular importance in our sample. As an example, about 80% of Dutch patents in our sample have no forward citations in the April 2009 version of the Patstat database, whereas that number is about 45% in Germany. In contrast, the information on family size is very comprehensive in the database, which is an additional reason to use this indicator.

Measures of similarity between the three value indicators are reported in Table 2. The first three rows present the correlation coefficients between the three value indicators. Somewhat surprisingly, the first (SF) and the third ($GEOG$) value indicators are the most correlated. Overall, however, all the indicators are strongly correlated with each other. The share of patents for which the family size differs from one indicator to another is reported in the next three rows. For instance, the correlation coefficient between indicator SF and $WIDE$ computed on Belgian patents is 0.52 and 20% of the patents have a different value indicator. Interestingly, the first and the second value indicators have a large share of observations in common, but the correlation coefficients between the two indicators are the lowest.

4.3 Sample

The initial sample is composed of all the OECD countries. Yet, only a few industries in some countries satisfy the requirements in terms of the number of patent applications and the information available in order to be included in the analysis. Three successive filters were applied to the data, reducing the final sample to 6 industries in 4 countries. Clearly, we put the focus on the quality of the information retrieved rather than on the quantity. First, we impose that an industry has a minimum of 100 patents over the period 1988 to 1992 to ensure that the data are representative enough. This rules out small countries such as Greece, Luxembourg or Portugal. Second, there must not be too many patents with missing IPC codes in order to have a reliable allocation of patents to industries. Countries for which more than 10 percent of patents cannot be allocated are excluded from the sample. Many countries were excluded by using this filter including Finland, Italy, Norway, the UK and the US. Finally, data on R&D expenditures at the industry level also needs to be available. Among the remain-

Table 2: Measures of similarity between value indicators.

	BE	DE	FR	NL
Correlation coefficients, all observations				
(SF,$WIDE$)	0.52	0.56	0.47	0.67
(SF,$GEOG$)	0.82	0.90	0.92	0.87
($WIDE$,$GEOG$)	0.71	0.62	0.53	0.78
Share of distinct observations				
(SF,$WIDE$)	0.20	0.13	0.09	0.17
(SF,$GEOG$)	0.51	0.38	0.35	0.51
($WIDE$,$GEOG$)	0.54	0.41	0.36	0.53
Number of patents	5,750	142,714	58,756	13,112

Notes: The first three rows indicate the correlation coefficients between the three value indicator. The next three rows indicate the share of patents whose family size differs from one indicator to another. The last row indicates the number of priority filings allocated to each country for the period 1988 to 1992. Sources: PATSTAT 0409. Own computation.

ing countries, Austria and Switzerland are filtered out by this third criterion. In total, the 6 following industries are retained: *Chemicals and chemical products* (CHEM), *Pharmaceuticals* (PHAR), *Machinery and equipment* (MACH), *Office, accounting and computing machinery* (COMP), *Electrical machinery and apparatus* (ELEC), *Radio, TV and communication equipment* (COMM), for four countries: Belgium, Germany, France and the Netherlands. These countries are very similar to each other. They are geographically close, are at a similar level of development and have patent systems that are roughly comparable. The sample is thus very homogeneous, which increases our confidence in the validity of the model's assumptions. The data are extracted for five years, from 1988 to 1992. This period is chosen for two reasons. First, we need to account for a certain lag before the value indicator becomes available. Second, two studies from the early nineties present survey data on the propensity to patent of European companies. These data will be used to validate the present results.

Table 3 presents the average number of priority filings over the period 1988 to 1992, in both absolute level and relative to R&D expenditures. Variations in patenting rates in the order of 500% can be observed across both countries and across industries. The number of priority filings per unit of R&D expenditure in the Belgian electrical machinery and apparatus industry (ELEC), the German machinery and equipment industry (MACH) and the French radio, TV and communication equipment (COMM) are particularly high in comparison with the other countries in the same industry. These extreme values, however, must be taken with a pinch of salt. Errors in the measurement of R&D expenditures as well as misallocations of patents to industry may induce strong biases, witnessing the limitation of the use of patent data at the industry level. However, it will become apparent in the next section that the methodology put forward in this paper allows the estimation of the propensity component in a way that is somehow robust to these measurement errors.

Table 3: Priority filings, average over 1988 to 1992.

	Priority filings				**Relative to R&D**			
	BE	DE	FR	NL	BE	DE	FR	NL
CHEM	206	3,782	1,410	374	0.34	0.88	0.92	0.41
PHAR	104	1,786	789	191	0.21	0.40	0.54	0.33
MACH	141	4,526	1,733	393	0.47	2.65	1.34	1.48
COMP	68	2,133	926	237	0.48	0.67	1.28	2.42
ELEC	25	978	394	78	2.83	0.68	0.60	0.61
COMM	108	3,537	1,645	376	0.82	1.13	3.27	0.53

Notes: See main text for the description of industries. Patents have been assigned according to the first value indicator. Sources: PATSTAT 0409 and OECD STAN database. Own computation.

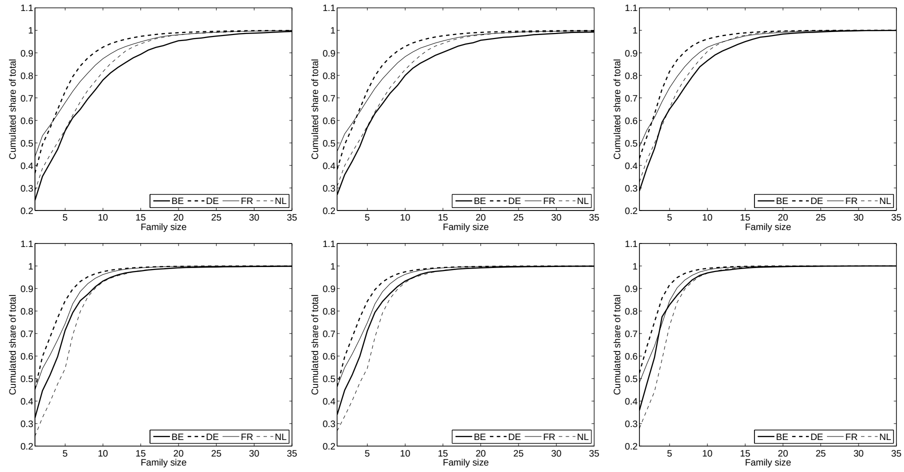
5 Results

5.1 Descriptive statistics

We are interested in understanding the extent to which differences in patenting rates can be explained by productivity or propensity effects. As explained in Section 3 a key element to look at is the distribution of patent value. A first insight into the value of patent applications is depicted in Figure 1. It plots the cumulative distribution of patent value as captured by the three indicators of family size for two distinct industries: chemicals and chemical products (CHEM, upper panels) and office, accounting and computing machinery (COMP, lower panels). The higher the curve, the higher the share of low value patents filed. For instance, between 70 and 80 percent of the patents filed in the German chemical industry have a family size lower than or equal to 5, as opposed to between 50 and 60 percent in Belgium. These statistics allow having a first idea of the relative propensities across countries. The model implies, indeed, that countries with a high share of low value patents have a high propensity to patent.

It is apparent from Figure 1 that patent practices differ across industries. Between 56 and 82 percent of the patents in the chemical industry have a family size less than or equal to 5, whereas the share for the computing machinery industry is between 54 and 92 percent. The difference is more pronounced for the share of patents with family size lower than or equal to 10, which averages at 0.88 and 0.96 in the chemical and the computing machinery industries respectively. Another observation that can be made from Figure 1 concerns the speed of convergence of the curves, which is slightly faster for the third value indicator (*GEOG*) than for the first two. This result is expected, given that

Figure 1: Cumulated distribution of patent value, 2 industries.



Notes: Upper and lower panels present the distribution of patent value for the chemicals industry (CHEM) and the office, accounting and computing machinery industry (COMP), respectively. Left (middle, right) panels are based on the SF ($WIDE$, $GEOG$) value indicator.

the family size of the third value indicator is bounded by the number of countries, whereas the family size measured in terms of patents is not bounded. The first two indicators have information in the tail that could prove useful for the identification procedure. Note, however, that the country curves exhibit a very similar pattern across indicators.

Table 4 presents the mean family size for the various value indicators. The mean size of the first value indicator (SF) exhibits both country and industry variations, though a similar pattern can be observed across countries. The pharmaceutical industry always has the largest family, the chemical industry comes second while the machinery and equipment industry generally comes last. The mean number of patents in the INPADOC family ($WIDE$, the second value indicator) behaves in a similar way as the first value indicator. Since the $WIDE$ family is composed of both priority filings and second filings, the family size may be driven by the high propensity to patent many minor inventions (or to file many patents for one invention) or by a high propensity to seek international protection with second filings. For instance, it is known that Japanese patents are more restrictive in scope than U.S. or European patents (they usually contain fewer claims). As a result, Japanese companies file more patents, and patent rates are considerably higher than in the U.S. or Europe. This is to be contrasted with the finding by Dernis et al. (2001) according to whom second filings citing multiple priority applications are particularly common for Japanese patents, "*which often cite between five and 30 priority applications for a single European or US patent.*" Our sample is relatively homogeneous across countries. The mean share of priority filings in the family is between 0.40 and 0.50 for the chemical and the pharmaceutical industries, and between 0.50 and 0.60 for the other industries. The number of countries in the INPADOC family is the third value indicator ($GEOG$). The information conveyed is close to that conveyed by the $WIDE$ indicator: larger families also cover more geographical markets. Yet, small discrepancies can be observed. For instance, the French radio, TV and communication equipment industry has a larger family than the electrical machinery and apparatus industry, but patents in the family protect fewer markets. By and large, however, one can conclude that the three value measures are relatively similar to each other. It is consistent with the strong correlation observed across indicators in Table 2.

Table 4: Descriptive statistics of value indicators.

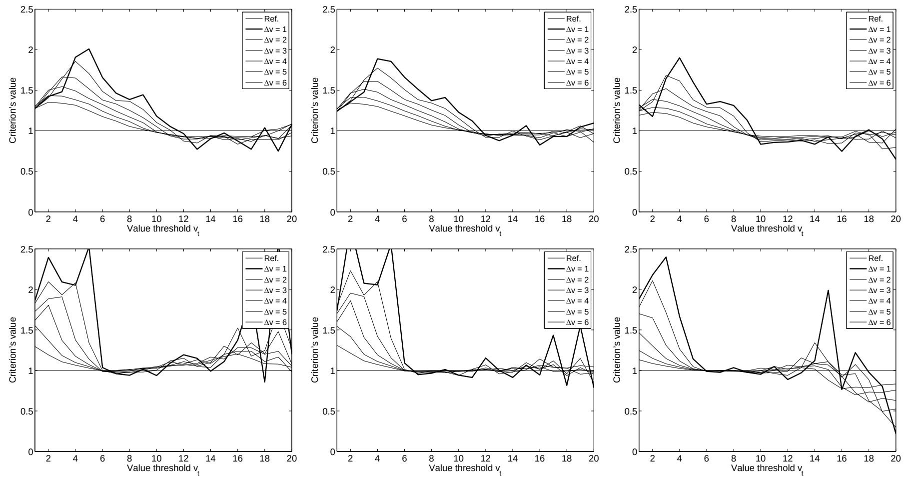
	BE	DE	FR	NL
Mean family size of indicator SF				
CHEM	6.90	4.19	4.74	5.95
PHAR	8.84	5.52	6.50	7.24
MACH	3.88	3.01	3.14	3.92
COMP	4.40	3.01	3.55	5.12
ELEC	4.39	2.91	3.37	4.76
COMM	4.69	2.99	3.55	5.09
Mean family size of indicator $WIDE$				
CHEM	6.75	4.14	4.60	5.88
PHAR	8.32	5.28	6.24	7.10
MACH	3.82	3.05	3.11	3.88
COMP	4.32	3.02	3.52	5.12
ELEC	4.28	2.92	3.34	4.76
COMM	4.61	2.99	3.52	5.09
Mean family size of indicator $GEOG$				
CHEM	5.27	3.38	3.90	4.67
PHAR	6.35	4.28	5.10	5.51
MACH	3.21	2.49	2.71	3.20
COMP	3.44	2.47	3.01	4.09
ELEC	3.54	2.39	2.89	3.89
COMM	3.66	2.45	2.30	4.09

Sources: Patstat 0409. Own computation.

5.2 Productivity and propensity

As shown in Section 3, it is possible to identify the *relative* productivity and propensity effects under two key assumptions. To begin with one reference country needs to be chosen, and the remaining countries will be evaluated against this benchmark. Germany is taken as the reference. In order to estimate the

Figure 2: Identification criterion (eq. 11) for two industries in the Netherlands as a function of v_t and Δv, for the 3 value indicators.



Notes: Upper and lower panels present the ratio of equation (11) for the chemical industry (CHEM) and the office, accounting and computing machinery industry (COMP). Left (middle, right) panels are based on the first (second, third) value indicator.

propensity and the productivity components, the appropriate value threshold v_t has to be determined. According to the identification criterion derived in Section 3, v_t must be chosen such that equation (11) holds. Remember that v_t is the threshold of invention value above which applicant from different countries roughly exhibit the same patenting behavior. Figure 2 presents the value of equation (11) on the y-axis evaluated at various threshold v_t (depicted on the x-axis), for 6 different steps ($\Delta v = 1, ..., 6$). The three upper panels present the series for the Dutch chemical industry while the lower panels present the series for the Dutch computing machinery industry. Panels on the left, the middle and the right rely on the first (SF), the second ($WIDE$) and the third ($GEOG$) value indicator, respectively.

As an example, the ratio for the computing machinery industry evaluated, say, with the $WIDE$ value indicator at a threshold value of 12 (that is, a family size of 12) and a step of 2 (that is, for family sizes of 12 and 13) is computed as follows (it corresponds to the middle panel in the lower part of Figure 2): Take the number of Dutch patents with a family size higher than or equal to 12 (56.198), multiplied by the number of German patents with value 12 or 13 (56.198*77.995 = 4,383.163), divided by the product of the number of German patents with value greater than or equal to 12 and the number of Dutch patents with value 12 or 13 (4,383.163/(195.525*20.969) = 1.069). The ratio obtained is close to 1, suggesting that the propensity to patent of the Dutch machinery and equipment industry relative to that of Germany can be evaluated at a threshold value of 12.

The condition is nevertheless satisfied for lower value threshold as well, typically above 5 in this case. The value threshold is higher in the chemical industry, where the ratio $\dot{P}_{NL}\ddot{P}_{DE}/\dot{P}_{DE}\ddot{P}_{NL}$ oscillates around 1 for families larger than 10. The fact that the series behave as expected (i.e., the ratio tends to 1 as the value threshold increases) is remarkable and increase our confidence in the model and its underlying assumptions, as well as in the value indicator that is used.

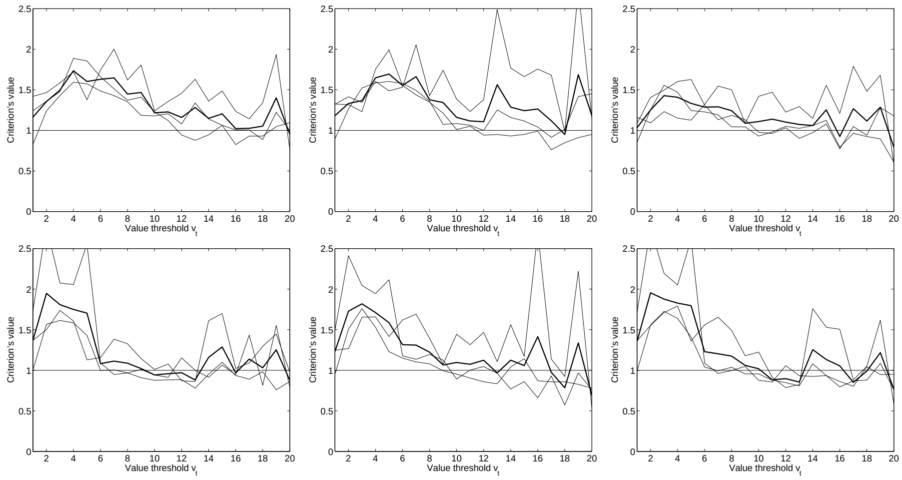
An important remark must be made regarding the variance that seems to increase with family size. (This phenomenon is well illustrated with the computing machinery industry, particularly with the second value indicator.) Since high value patents are less frequent than low value ones, the weight given to each individual patent increases with the value threshold such that one single patent can be responsible for a large deviation in the ratio. For instance, there are only 0.53% and 0.20% of patents by Dutch and German inventors that have a family equal to 14 (corresponding to 5.52 and 20.22 patents respectively). If one more "highly valuable" patent would be attributed to the Netherlands, the ratio would increase by 14% from 0.93 to 1.06. An additional patent would bring the identification criterion to 1.18. In this case, two additional high value patents would thus be responsible for a 27% increase in the ratio. For this reason, the behavior of the series above a too high threshold should be disregarded. The effect of individual patents is somewhat mitigated when more steps are taken into account, but a criterion computed with a too large Δv is subject to the statistical artifact described in Section 3.

Two conditions must be met for the identification to be valid. First, equation (11) evaluated at v_t must be equal to 1 for each country. Second, v_t must be similar across countries. A summarized view for all the industries is given by Figure 3, where the bold curve in each panel represents the mean value of the identification criterion across the 3 countries (BE, FR, NL) for the $WIDE$ indicator. Equation (11) seems to hold for threshold values between 10 and 20, depending on the industry. (Note that the criterion has been computed with a step of 1 ($\Delta v = 1$), the most conservative specification.) The ratio of propensities in equation (7) will thus be computed for value thresholds ranging from 10 to 20. In so doing, we will also have an idea of the sensibility of the indicator to the chosen threshold.

The propensity component is presented in Figure 4 for the chemical and the computing machinery industries (upper and lower panels, respectively), for each of the three value indicators.

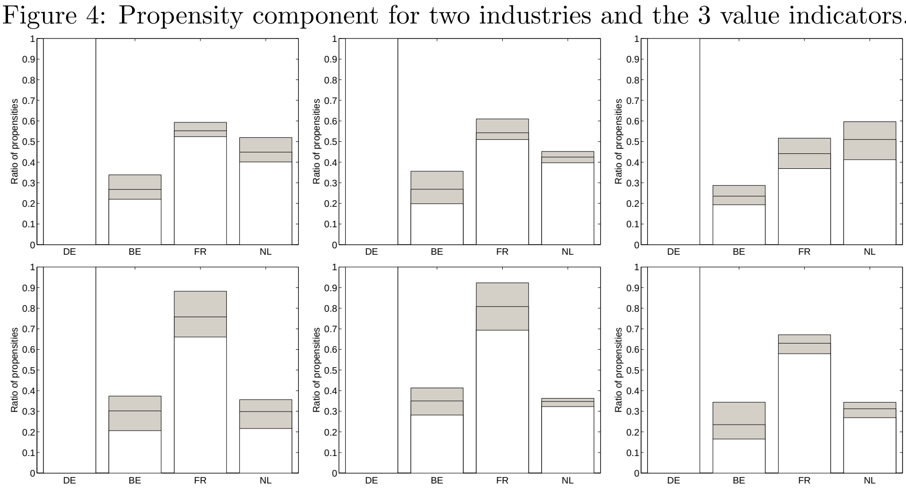
In the chemical industry, for instance, Germany always has the highest propensity to patent (which is 1, by construction, since Germany is the reference country), Belgium has a propensity that is a quarter to a third that of Germany and France and the Netherlands are about half the level of Germany. The grey area over each bar is a confidence interval. It corresponds to the ratios obtained for value thresholds ranging from 10 to 20. The bar in the middle of the grey area corresponds to the average ratio obtained. Overall, the confidence interval is reasonably small.

Figure 3: Summary of identification criterion (eq. 11) for each of the 6 industries, second value indicator ($WIDE$).



Notes: From top-left to bottom-right: CHEM, PHAR, MACH, COMP, ELEC and COMM. The identification criterion has been computed with a step of 1 ($\Delta v = 1$). The bold line represents the mean across countries.

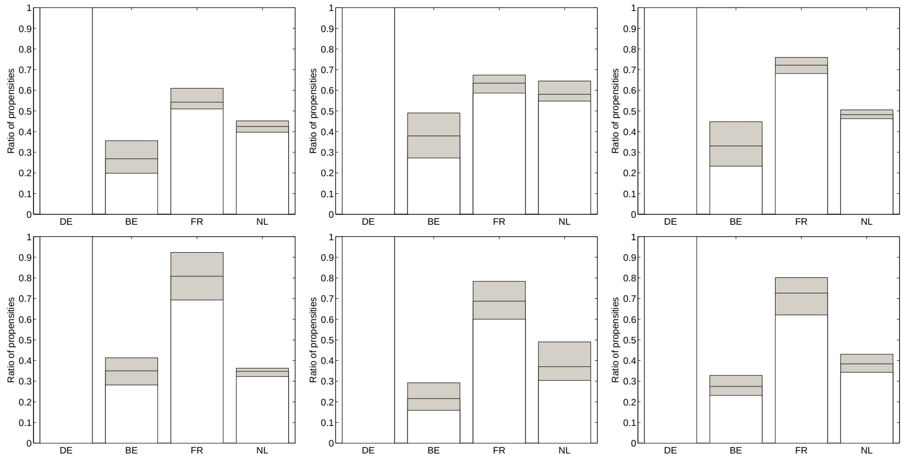
Figure 4: Propensity component for two industries and the 3 value indicators.



Notes: Upper and lower panels present the propensity component for the chemical industry (CHEM) and the office, accounting and computing machinery industry (COMP) respectively. Left (middle, right) panels are based on the SF ($WIDE$, $GEOG$) value indicator.

The propensity components are quite robust to changes in the value indicator that is used. We choose to estimate the propensities for the remaining industries with the second value indicator ($WIDE$). It is more appropriate than the first one (SF) as it corrects for the double counting of second filings and has more heterogeneity in the tail than the third one ($GEOG$). Figure 5 depicts the propensity component for the 6 industries computed with the second value indicator. Germany always has the highest propensity to patent, though the French office, accounting and computing machinery industry (COMP) almost matches the level of Germany. The propensity is always the lowest in Belgium, oscillating around 20 to 40 percent of the level of Germany, and the Netherlands is situated between France and Belgium. The propensity rates are easy to interpret. Assume that the German radio, TV and communication equipment industry (COMM) has 100 inventions and file 100 (50) patent applications. With the same set of inventions, the French would file about 70 (35) patents, the Dutch 40 (20) and the Belgian 30 (15).

Figure 5: Propensity component for the second value indicators, all industries.



Notes: From top-left to bottom-right: CHEM, PHAR, MACH, COMP, ELEC and COMM.

Table 5 presents the key results of the identification. Column (A) shows the patent-to-R&D rates computed with the $WIDE$ indicator.[7] In column (B), the patenting rates are expressed relative to Germany. For instance, the French chemical industry has 6% more patents per R&D than the German industry has. These relative differences are split into a propensity and a productivity effect, reported in columns (C) and (D) respectively. Column (C) summarizes the information presented in Figure 5. Following equation (6), the elements in column (D) are obtained by dividing the elements in column (B) by the relative propensity component of column (C). Interestingly, the relative patenting rates in column (A) are not always reflective of differences in the productivity of

[7] The difference between column (A) of Table 5 and Table 3 are due to the fact that patents are counted using the first value indicator (SF) in Table 3.

research. For instance, the German chemical industry has 3 times more patents per unit of R&D than the Belgian industry has, although Belgium is found to be slightly more productive than Germany. Similarly, the French and the German chemical industries have about the same number of patents, but France is found to be more productive. Relative productivity levels are roughly between 0.50 and 2.50 times the level of Germany although some extreme values can be observed (such as for the Belgian electrical machinery and apparatus industry). The results are discussed in the next section.

5.3 Interpretation and discussion of the results

We start by explaining why extreme values are observed for the productivity component. We then assess whether the propensity rates computed are reasonable and bring a tentative explanation for the differences in the propensity observed across countries.

The productivity levels for three country-industry pairs are excessively high: NL-COMP, BE-ELEC and FR-COMM. The presence of these outliers is due to the mismeasurement and misallocation of R&D expenditures and patent data. Since the propensity component is estimated solely from the distribution of patent value it is robust to potential measurement errors, and the productivity component absorbs most of the errors. In this respect, it is symptomatic that the outliers are observed for country-industry pairs that already have an abnormally high patent-to-R&D ratio (see column (A) of Table 5). A way to circumvent this problem would be to perform the analysis at the country level, where the measurement errors are less severe and the misallocation problem becomes irrelevant.

It is challenging to assess the accuracy of the propensity rates obtained given that the propensity effect is largely unobserved. We have to turn to indirect ways to evaluate how realistic the results are. All in all, the body of evidence tends to show that the propensity rates recovered are in a reasonable range.

A first way to assess whether the results obtained are coherent consists in estimating the correlation between columns (B) and (C) of Table 5. While column (B) is a measure of the patenting rates relative to Germany, column (C) is a measure of the relative propensities to patent. Obviously, we expect these two measures to be somehow correlated even though differences in productivity levels also account for differences in patenting rates. The correlation coefficients are computed for two groups of industries: PHAR and CHEM, and MACH, COMP, ELEC and COMM. They are equal to 0.66 and 0.50 respectively and are significant at the 10% probability threshold.[8] This significant correlation coefficient is remarkable since the propensity component in column (C) does not use information on the research efforts but solely relies on the information contained in the distribution of patent value. It is a first sign that some useful information has been extracted.

[8] Three dubious observations have been excluded from the second group of industries: NL-COMP, BE-ELEC and FR-COMM. Note that the results also hold when the reference country is excluded.

Table 5: Overview of productivity and propensity effects.

	(A) PF/R&D	(B) $(A)_i/(A)_{DE}$	(C) Relative prop.	(D) Relative prod.
Chemicals and chemical products (CHEM)				
BE	0.29	0.36	0.27	1.33
DE	0.80	1.00	1.00	1.00
FR	0.85	1.06	0.54	1.95
NL	0.37	0.43	0.42	1.07
Pharmaceuticals (PHAR)				
BE	0.16	0.45	0.38	1.19
DE	0.35	1.00	1.00	1.00
FR	0.48	1.35	0.63	2.12
NL	0.28	0.58	0.58	1.35
Machinery and equipment (MACH)				
BE	0.44	0.18	0.33	0.54
DE	2.49	1.00	1.00	1.00
FR	1.29	0.52	0.72	0.72
NL	1.39	1.08	0.48	1.16
Office, accounting and computing machinery (COMP)				
BE	0.45	0.71	0.35	2.02
DE	0.63	1.00	1.00	1.00
FR	1.22	1.94	0.81	2.40
NL	2.14	1.75	0.35	9.75
Electrical machinery and apparatus (ELEC)				
BE	2.62	4.11	0.22	19.05
DE	0.64	1.00	1.00	1.00
FR	0.59	0.92	0.69	1.33
NL	0.56	0.96	0.37	2.37
Radio, TV and communication equipment (COMM)				
BE	0.76	0.71	0.27	2.60
DE	1.06	1.00	1.00	1.00
FR	3.12	2.93	0.73	4.04
NL	0.47	0.15	0.38	1.16

Notes: column (A) reports the average number of priority filings per thousands U.S. dollars PPPs of R&D expenditures over the period 1988 to 1992. Sources: Patstat 0409 and OECD MSTI database. Own computation.

A second way consists in checking whether the results are consistent with existing evidence on the propensity to patent. Two studies surveyed patent practices of European firms during the early nineties. Using data from a survey of Europe's largest R&D-performing industrial firms, Arundel and Kabla (1998, p. 136) estimate that German firms are significantly more likely than other European firms to patent a larger share of their product innovations. This is consistent with the high propensity rates obtained for Germany for all industries. The second survey has been conducted by the EPO in 1994. The report contains country-level descriptive statistics on the proportion of applications filed for patentable inventions. It is difficult to extract exact numbers for the data presented, but it clearly appears that Germany has the highest propensity to patent, followed by France and the Netherlands (EPO, 1994, p. 107). Somewhat surprisingly is the high score reported for Belgium, which ranks first with Germany. This observation must be taken with a pinch of salt: the data presented in the study are (presumably) not weighted by company size such that the mean rates may not be representative of the true country-average.[9]

Note that the propensity rates are also consistent with the differences in the cost of obtaining a patent. Given the significant sensibility of applicants to patent fees, one would expect a higher propensity to patent in countries where patent protection is cheap. In a study of patent fees in 1992, Helfgott (1993) reports that fees in Germany are the lowest at $386, and that they are of $865 in France $1,262 in the Netherlands. Fees in Belgium were not computed. However, a study by de Rassenfosse and van Pottelsberghe (2007) shows that patent fees in Belgium were among the most expensive in Europe in 2003. If no major change in fees occurred during the nineties, one can logically assume that they were also among the highest at the beginning of the nineties.

It is striking to note the relative stability of the propensity component across industries (e.g., Belgium has always around one-third of the propensity to patent of Germany), suggesting a strong country effect. As a matter of fact, the incentives to apply for patents are mostly driven by national factors such as patent fees, the strength/quality of the patent system, the IP awareness, etc. One must not forget that the propensity rates presented in Table 5 are free of any industry effect. For instance, the Belgian computing industry is compared with the German computing industry such that industry-specific factors are wiped out.

One can wonder why such differences in the propensity to patent are observed across countries that are otherwise very similar. We believe that these differences are mainly driven by the market size (not to mention differences in patent fees). The profitability of an invention rises with market size; more inventions are thus likely to be profitable enough to be patented in larger markets. An indirect support for this argument is provided by Harhoff et al. (2009). The authors look at the flow of international patent extension and find that companies are more willing to seek patent protection in larger markets. Note that the high propensity to patent of German firms could also be due to the

[9] In fact, this bias is very likely to affect the Belgian data. Propensity rates for Belgium (EPO, 1994, p. 107) are at odds with the number of patented inventions owned by Belgian companies (EPO, 1994, p. 114). Whereas mean propensity rates for Belgian companies are similar to German companies, 60% of the Belgian companies own fewer than 5 patents as opposed to 33% for Germany. In contrast, 66% of Dutch companies own fewer than 5 patents though the mean propensity rate in the Netherlands is well below that of Belgium. It corroborates the hypothesis that Belgian companies surveyed were probably relatively small and that the aggregate propensity rate is not representative of the overall propensity to patent of Belgian firms.

German Employees' Invention Act, which organizes the transfer of the right to an invention to the inventor if it is not claimed by the employee (Harhoff and Hoisl, 2007). In addition to raising IP awareness, the Act presumably increases firm's incentive to apply for patents.

6 Concluding remarks

It is well known that patents are imperfect indicators of inventive activity: Not all inventions are patentable and not all patentable inventions are actually patented. As a result, patent-to-R&D ratios reflect both a productivity effect that leads from research efforts to inventions and a propensity effect that leads from inventions to patents. The present paper proposes a solution to this identification problem in patent statistics.

We propose a methodology to evaluate *relative* propensity and productivity rates. It relies on a minimum number of plausible assumptions and can easily be applied to patent data. If a universal density function of invention value is assumed, it is shown that it is sufficient that the probability of patenting inventions of high enough value be similar across countries to identify the effects. The proposed framework also allows a better grasp of the ins and outs of current patent indicators. For instance, the count of triadic patents can be seen as a special case of the model. In order to facilitate the application of the model to real data, an easy-to-implement criterion to select the appropriate value threshold is derived. The method is then applied to a novel data set of priority patent applications in which each patent is fractionally allocated to its inventors' countries and to the technological areas to which it belongs. The analysis is performed for six industries in four countries over the period 1988 to 1992 and leads to results that are very intuitive. Interestingly, it is frequently found that more patents per R&D are not reflective of higher productivity of research. In some cases, indeed, countries with high patenting rates are found to be less productive than countries with lower patenting rates.

Because it facilitates the interpretation of patent statistics and allows to better monitor innovation performances, the methodology should be appealing to policymakers. It is also of interest to scholars as it provides a tool to improve our understanding of the invention process. For instance, existent studies on the determinants of innovation could be replicated using the productivity component estimated with the methodology put forward in this paper.

Some limitations of the model and the empirical application need to be briefly mentioned. The methodology is as valid as the underlying assumptions, and the results are as reliable as the data and the value indicator that are used. As far as the identification procedure is concerned, we have explained why the assumptions made are very general. However, if one does not believe that the propensity to patent is roughly equal across countries for inventions that are valuable enough, then the use of simple patent-to-R&D ratio is in any case severely biased. However that may be, it is likely that the propensity to patent is more similar across countries for high value inventions than for low value ones, thereby reducing the bias of patent indicators if not fully eliminating it. In this respect, it must be noted that the results of the empirical application corroborate the theory developed (in particular regarding the fact that the identification criterion is satisfied). If there exists no way to verify that the true propensity component has been identified, we are able to show that at least some useful information on propensity rates has been extracted as witnessed by the significant correlation between the R&D-patent ratio and the estimated propensity component. This correlation is remarkable given that the propensity component has been recovered solely from patent data, using no information on R&D expenditure. The high propensity to patent reported in the literature for Germany also clearly appears in the results. Regarding the empirical application, two limitations must be discussed. First, the value indicator chosen is potentially biased if patents are filed but not extended for, say, budgetary reasons or for lack of familiarity with international patenting. It is certainly true for countries that are drastically different from each other (such as countries that are at a different level of economic development or that have a different patent system), but we believe our sample is sufficiently homogeneous to avoid this shortcoming. In addition, the methodology requires only good information for high value inventions, where we expect financing constraints to be less severe than for inventions with a low potential. It does not matter if an invention that has a family of, say, 4 in one country would have a family of 8 in another if the chosen identification threshold is set at 10. Second, it has become apparent for some industries that the input or the output of the innovation process are misallocated. (Think of the Belgian electrical machinery and apparatus industry that has four to five times more patents per R&D expenditures than the German or the Dutch industry.) If the measure of the propensity is robust to such measurement errors, the productivity component, for its part, is very sensitive. A way to circumvent this problem would be to use additional information from patent data. Information on the number of inventors could be taken from the patent documents themselves instead of from external sources. This would, however, require significant advances in the harmonization of the name of inventors.

Various extensions to the current work can be envisaged. First, the focus of the present paper is on the quality of the information retrieved rather than on the quantity. Missing information on R&D expenditures and IPC classification of patents prevented us from extending the analysis to more countries and industries. A useful extension would thus be to compute the indicator for a larger sample. A second extension would be to study whether the methodology could be applied to firm-level or even to country-level data. Concerning the latter, our results suggest that the propensity component is relatively stable across industries (e.g., the propensity to patent in Belgium is systematically around one-third the German level) such that the industry composition of research activity should have only a limited impact on the aggregate propensity rates. A country-level analysis would also be more robust to mismeasurement and misallocation of R&D and patent data. The productivity component being particularly vulnerable to these issues, a country-wide study would provide a more reliable picture of research productivity. A third extension would be to make the results less dependent on a particular value indicator. Multiple indicators could be combined (such as citations and renewal data) to refine the measurement of patent value.

References

Arundel, A., Kabla, I., 1998. What percentage of innovations are patented? Empirical estimates for European firms. Research Policy, 27(2), 127-141.

Brouwer, E., Kleinknecht, A., 1997. Measuring the unmeasurable: A country's non-R&D expenditure on product and service innovation. Research Policy, 25(8), 1235-1242.

Cassiman, B., Veugelers, R., 2002. R&D cooperation and spillovers: Some empirical evidence from Belgium. The American Economic Review, 92(4), 1169-1184.

Cohen, W., Nelson, R., Walsh, J., 2000. Protecting their intellectual assets: Appropriability conditions and why U.S. manufacturing firms patent (or not). NBER Working Paper 7552.

Coombs, R., Narandren, P., Richards, A., 1996. A literature-based innovation output indicator. Research Policy, 25(3), 403-413.

Crépon, B., Duguet, E., Mairesse, J., 1998. Research, innovation and productivity: An econometric analysis at the firm level. Economics of Innovation and New Technology, 7(2), 115-158.

Czarnitzki, D., Toole, A., 2010. Patent protection, market uncertainty, and R&D investment, Review of Economics and Statistics, forthcoming.

Danguy, J., de Rassenfosse, G., van Pottelsberghe de la Potterie, B., 2009. The R&D-patent relationship: An industry perspective. EIB Papers, 14(1), 170-195.

de Rassenfosse, G., 2010. How much do we know about firms' propensity to patent and should we worry about it? Université libre de Bruxelles, ECARES Working paper, forthcoming.

de Rassenfosse, G., Dernis, H., Guellec, D., Picci, L., van Pottelserghe de la Potterie, B., 2010. A corrected count of priority filings. OECD Working Paper, forthcoming.

de Rassenfosse, G., van Pottelserghe de la Potterie, B., 2007. Per un pugno di dollari: A first look at the price elasticity of patents. Oxford Review of Economic Policy, 23(4), 588-604.

de Rassenfosse, G., van Pottelserghe de la Potterie, B., 2009. A policy insight into the R&D-Patent relationship. Research Policy, 38(5), 779-792.

Dernis, H., Guellec, D., van Pottelsberghe de la Potterie, 2001. Using patent counts for cross-country comparisons of technology output. OECD STI Review, 27, 129-146.

European Patent Office (EPO), 1994. Utilisation of patent protection in Europe. Eposcript, Volume 4. Munich: European Patent Office, 275 p.

Godin, B., 2005. Measurement and statistics on science and technology: 1920 to the present. Routledge Studies in the History of Science, Technology and Medicine. Oxfordshire: Routledge, 360 p.

Grabowski, H., Vernon, J., 1994. Returns to R&D on new drug introductions in the 1980s. Journal of Health Economics, 13(4), 383-406.

Griliches, Z., 1990. Patent statistics as economic indicators: A survey. Journal of Economic Literature, 28(4), 1661-1707.

Grupp, H., Schmoch, U., 1999. Patent statistics in the age of globalisation: New legal procedures, new analytical methods, new economic interpretation. Research Policy, 28(4), 377-396.

Harhoff, D., Hoisl, K., 2007. Institutionalized incentives for ingenuity. Patent value and the German Employees' Inventions Act. Research Policy, 36(8), 1143-1162.

Harhoff, D., Hoisl, K., Reichl, B., van Pottelsberghe de la Potterie, B., 2009a. Patent validation at the country level - The role of fees and translation costs. Research Policy, 38(9), 1423-1437.

Harhoff, D., Hoisl, K., Webb, C., 2008. European patent citations — How to count and how to interpret them? Mimeo, Ludwig-Maximilians-Universität, Munich.

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

Helfgott, S., 1993. Patent filing costs around the world. Journal of the Patent and Trademark Office Society, 75(7), 567-580.

Jaffe, A., Lerner, J., 2004. Innovation and its discontents: How our broken patent system is endangering innovation and progress, and what to do about it. New Jersey: Princeton University Press, 259 p.

Jaffe, A., Trajtenberg, M., 1996. Flows of knowledge from universities and federal laboratories: Modeling the flow of patent citations over time and across institutional and geographic boundaries. Proceedings of the National Academy of Sciences, 93, 12671-12677.

Jensen, P., Palangkaraya, A., Webster, E., 2006. Disharmony in international patent office decisions. The Federal Circuit Bar Journal, 15(4), 679-712.

Kleinknecht, A., Van Montfort, K., Brouwer, E., 2002. The non-trivial choice between innovation indicators. Economics of Innovation and New Technology, 11(2), 109-121.

Kortum, S., 1993. Equilibrium R&D and the Patent-R&D Ratio: U.S. Evidence. The American Economic Review, 83(2), 450-457.

Kortum, S., Putnam, J., 1997. Assigning patents to industries: Test of the Yale Technology Concordance. Economic Systems Research, 9(2), 161-176.

Lanjouw, J., Pakes, A., Putnam, J., 1998. How to count patents and value intellectual property: the uses of patent renewal and application data. The Journal of Industrial Economics, 46(4), 405-432.

Lanjouw, J., Schankerman, M., 2004. Patent quality and research productivity: Measuring innovation with multiple indicators. The Economic Journal, 114(495), 441-465.

Lerner, J., 1994. The importance of patent scope: An empirical analysis. The RAND Journal of Economics, 25(2), 319-333.

McDowell, J., 1982. Obsolescence of knowledge and career publication profiles: Some evidence of differences among fields in costs of interrupted careers. The American Economic Review, 72(4), 752-768.

Mejer, M., van Pottelsberghe de la Potterie, B., 2009. Economic incongruities in the European Patent System. Université libre de Bruxelles, ECARES Working Paper 2009-003.

Meyer, M., 2000. Patent citations in a novel field of technology — What can they tell about interactions between emerging communities of science and technology? Scientometrics, 48(2), 151-178.

Organisation for Economic Co-operation and Development (OECD), 2009. OECD Patent Statistics Manual. Paris: OECD, 158 p.

Pakes, A., 1986. Patents as options: Some estimates of the value of holding European patent stocks. Econometrica, 54(4), 755-784.

Patel, P., Pavitt, K., 1995. Patterns of technological activity: their measurement and interpretation. In P. Stoneman (Ed.), Handbook of the economics of innovation and technological change, Oxford: Blackwell, 608 p.

Pavitt, K., 1985. Patent statistics as indicators of innovative activities: Possibilities and problems. Scientometrics, 7(1-2), 77-99.

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

Scherer, F., Harhoff, D., 2000. Technology policy for a world of skew-distributed outcomes. Research Policy, 29(4-5), 559-566.

Schmoch, U., Laville, F., Patel, P., Frietsch, R., 2003. Linking technology areas to industrial sectors. Final report to the European Commission, DG Research. Karlsruhe (DE), Paris (FR), Brighton (UK), 77 p.

Schmookler, J., 1954. The level of inventive activity. The Review of Economics and Statistics, 36(2), 183-190.

Soete, L., Wyatt, S., 1983. The use of foreign patenting as an internationally comparable science and technology output indicator. Scientometrics, 5(1), 31-54.

Tong, X., Frame, J., 1994. Measuring national technological performance with patent claims data. Research Policy, 23(2), 133-141.

Trajtenberg, M., 1990. A penny for your quotes: Patent citations and the value of innovations. The RAND Journal of Economics, 21(1), 172-187.

van Pottelsberghe, B., van Zeebroeck, N., 2008. A brief history of space and time: The scope-year index as a patent value indicator based on families and renewals. Scientometrics, 75(2), 319-338.

van Zeebroeck, N., 2010. The puzzle of patent value indicators. Economics of Innovation and New Technology, forthcoming.

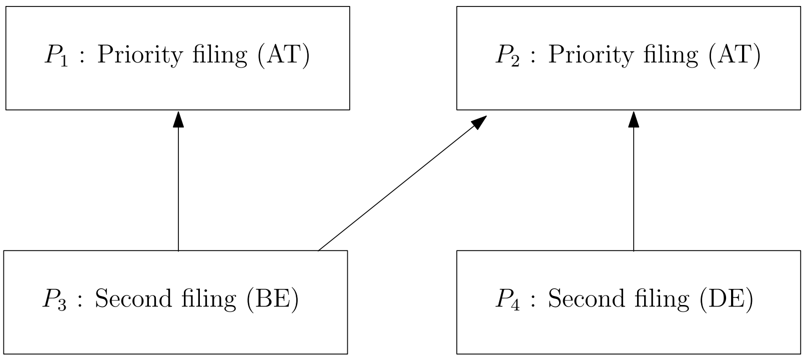
Verspagen, B., van Moergastel, T., Slabbers, M., 1994. MERIT concordance table: IPC-ISIC. MERIT Research Memorandum 2/94-004.

World Intellectual Property Organization (WIPO), 2008. World Patent Report. Geneva: WIPO, 70 p.

A Description of value indicators

This appendix presents an illustrative example of how patent family size is computed. Let us assume that the family size is computed for 2 Austrian priority filings P_1 and P_2. As indicated in Figure 6, patent P_3, filed in Belgium, claims priority of both P_1 and P_2. P_4, filed in Germany, claims priority of P_2 only.

Figure 6: An illustrative example of patent family.



Notes: The country code in parentheses indicates the patent office of application.

Table 6 gives information on how the score for each patent is computed. According to the first value indicator, P_1 has a family of size 2 and P_2, which has 2 second filings, has a family of size 3. The second value indicator considers that the 4 patents belong to the same family (P_1 and P_2 are linked through P_3). The family is thus of size 4 and 50% of the size (i.e., value) is attributed to each priority filing. The family of the third indicator ($GEOG$) is the same as that of the $WIDE$ indicator. However, since the family covers only 3 markets, its size is of 3.

Table 6: Score of the value indicators.

	Indicator SF	**Indicator** $WIDE$	**Indicator** $GEOG$
Family of P_1	$\{P_1, P_3\}$	$\{P_1, P_2, P_3, P_4\}$	$\{P_1, P_2, P_3, P_4\}$
Family of P_2	$\{P_2, P_3, P_4\}$	$\{P_1, P_2, P_3, P_4\}$	$\{P_1, P_2, P_3, P_4\}$
Size of family P1	2	4	3
Size of family P2	3	4	3
Score given to P1	2	2 (0.5*4)	1.5 (0.5*3)
Score given to P2	3	2 (0.5*4)	1.5 (0.5*3)