

Aspects and availability of patent data

Patent data provide rich information on technical characteristics of inventions, development and ownership of inventions and the history of applications. The amount of available patent data is large, covering all countries with patent systems, often with information at a detailed geographic level and over longer time periods than most other datasets. However, the use of patent data to assess innovation performance requires awareness of several caveats, such as the fact that not all inventions are patented, patenting differences across fields of activity, the substantial variation in the value of patents and the fact that patenting will be critically affected by different legal IP systems, all of which do not allow simple comparisons across countries. If these factors are adequately taken into account, patent data can be used as measures of inputs to innovation, indicators of economic performance, as well as measures of characteristics of innovative activities. They can also enable an assessment of the economic value of inventions and evaluation of aspects of the patent system itself.

What type of information does patent data provide?

Patent data provide information on the following characteristics of an invention or innovation:

- **Technical characteristics of the invention.** Characteristics include a description of the invention, the list of claims that defines the scope of protection of the patent rights (legal boundaries), the technology fields to which the invention pertains, patent references that are citations to previous patents relevant for the invention, and non-patent references (e.g. scientific publications, conference proceedings, books, database guides) to which the invention relates.
- **Development and ownership of the invention.** This information includes, the inventors' and applicants' names and addresses. The applicants are those who will own the patent if it is granted.
- **History of the application.** This information includes, for instance, the first date of filing of a patent application; the date of grant, refusal or withdrawal; and the first day that protection will apply in the country.

Patent data can be analyzed in diverse ways, differing in terms of the level of aggregation of the data compiled (national, regional, company level, industry or technical field level); the approach taken (compilation of indicators, performance of econometric estimates); and the analytical or policy questions addressed. Examples of the information that can be gleaned from patent data include the following:

- **Input to innovation.** While patents can be seen as R&D output, they can also be seen as input to innovation (e.g. when the invention is used downstream in economic processes). This intermediate character makes patent data a useful bridge between R&D data and innovation data.
- **Economic performance.** In a study of 258 R&D professionals, Keller and Holland (1982) concluded that the number of an inventor's patents is significantly correlated with superior performance ratings and self-rating. In a study of 1,200 companies in high-technology industries, Hagedoorn and Cloudt (2003) concluded that the number of patents filed by a company is a very good reflection of its technological performance. At the country level, de Rassenfosse and van Pottelsberghe (2008) found a high correlation between patent numbers

and R&D performance.

- **Characteristics of innovative activities.** (see [Characteristics of innovative activities](#) [1]).
- **The economic value of inventions.** (see [Patent valuation methods](#) [2]).
- **The patent system itself.** These might be characteristics, for example, based on the volume of patenting activity by companies or the way patent offices operate.

What is the availability of patent data?

- **The amount of patent data available for researchers is substantial.** More than one million patents are applied for worldwide each year, providing unique information on the progress of invention. Patent data are public, unlike survey data which are usually protected by statistical secrecy laws.
- **The spatial and temporal coverage of patent data is large.** Patent data are available from all countries with a patent system, i.e. nearly all of the world's countries. They are available—sometimes in electronic form—from first patent systems, which go back to the 19th century in most OECD countries.
- **The range of technologies covered by patent data is wide.** Patents provide information on technologies for which there are sometimes few other sources of data (e.g. nanotechnology).
- **Patent data are readily available from national and regional patent offices.** The marginal cost for the statistician is much lower than for conducting surveys, although it is sometimes still significant (data need to be cleaned, formatted, etc.). Unlike survey data, collection of patent statistics does not put any supplementary burden on the reporting unit (e.g. business) because the data are already collected by patent offices in order to process applications.

Although patent data are produced by the patent authorities, patent databases using such data are also produced and published by private entities. Some patent databases widely used for statistical and research purposes are:

- the NBER Patent Citations Data Files created by Jaffe, Trajtenberg and Hall, with the assistance of researchers at the NBER and Case Western Reserve University;
- the EPO Worldwide Patent Statistical Database (also known as EPO PATSTAT) created by the EPO with the OECD Patent Statistics Task Force; and
- the IIP (Institute of Intellectual Property) patent database, which gathers internal patent data from JPO (Seiri Hyojunka Data).

What are the caveats to using patent data as measures of inventive activities?

Patent data have the following limitations in reflecting inventive activities, which need to be taken

into account if they are used for analysis:

- **Not all inventions are patented.** Inventions with few economic possibilities may not justify the cost of patenting. Inventions that make a trivial contribution to the art and non-technological inventions do not qualify under the legal requirements of patenting. Strategic considerations may lead the inventor to prefer alternative protection (secrecy), with the result that the patent data do not reflect such inventions (e.g. Pavitt, 1988).
- **Different industry levels of filing.** The propensity to file patent applications differs significantly across technical fields. For instance, in the electronics industry (e.g. semiconductors) a patented invention can be surrounded by patent applications on incremental variations of the invention with a view to deterring the entry of new competitors and to negotiating advantageous cross-licensing deals with competitors. As a result of this “patent flooding” strategy, some technical fields have a larger number of patents than others.
- **Companies’ propensity to patent also differs.** New or small and medium-sized enterprises (SMEs)—notably those that lack large-scale production—have more difficulty covering the costs of a patent, although national policies attempt to deal with this problem by providing SMEs with subsidies or discount rates.
- **The value of patents varies greatly.** The distribution of patents is highly skewed (e.g. Pakes and Schankerman, 1986). Many patents have no industrial application, hence are of little or no value to society, whereas a few have very high value. With such heterogeneity, simple patent counts can be misleading. Moreover, according to the PatVal survey (2005), about 40% of patents in the sample are not used for industrial or commercial purposes for strategic reasons or because the owners lack the complementary downstream assets to exploit them: 18.7% are not used and aim to block competitors and 17.4% are considered “sleeping patents” that are not used at all.
- **Patent laws differ.** Differences in patent law and practices around the world limit the comparability of patent statistics across countries; therefore, it is preferable to use homogenous patent data (coming from a single patent office or single set of patent offices).
- **Legal changes confound analyses.** Changes in patent laws over the years call for caution when analyzing trends over time. The protection afforded patentees worldwide has been increased since the early 1980s, and companies are therefore more inclined to patent than before. The list of technologies covered has grown longer over time and in some countries now includes software and genetic sequences, which were previously excluded. Other variables, such as office administration, can have a substantial impact on patent counts, notably patents granted, during a particular time period.
- **Patent data are complex and difficult to interpret.** Patent data are complex to interpret as they are generated by complex legal and economic processes. Failure to take into account all the surrounding factors when compiling and interpreting patent data can lead to erroneous conclusions.
- **Time lag in data availability and unavailable data.** No statistics are available until 18 months after the priority date, since the application is not published until then. This limits the legally possible timeliness of patent data. Patent offices publish aggregate counts of recent applications for the purpose of monitoring their own activity, but these data are not accessible to outside users and cannot be exploited for analytical purposes.

Most of these limitations can be overcome using appropriate methodologies to address data bias and limitations in order to limit their impact. For example, the issue of the skewed distribution of patent value can be addressed by weighting patent counts by number of citations, or by selecting a sub-

sample of patents that are of similar value (e.g. triadic patents which typically capture high-value patents). Similarly, to surmount the drawbacks associated with differing propensities to patent across industries, the analysis can be restricted to a sector or industry or the data weight adjusted appropriately.

References

- de Rassenfosse, G. and B. van Pottelsberghe (2008), “A policy insight into the R&D patent relationship”, ULB Working Paper.
- Hagedoorn, J. and M. Cloudt (2003), “Measuring innovative performance: Is there an advantage in using multiple indicators?” *Research Policy*, Vol. 32, pp. 1365–79.
- Keller, R. T. and W. E. Holland (1982), “The measurement of performance among R&D professional employees: A longitudinal analysis”, *IEEE Transactions of Engineering Management*, No. 29, pp. 54–58.
- OECD (2009), *OECD Patent Statistics Manual*, OECD Publishing, Paris. doi: 10.1787/9789264056442-en
- Pakes, A. and M. Schankerman (1986), “Estimates of the value of patent rights in European countries during the post-1950 period”, *Economic Journal*, Vol. 96/December, pp. 1052–76.
- PatVal-EU (2005), “The value of European patents: Evidence from a survey of European inventors”, Final report of the PatVal-EU project.
- Pavitt, K. (1988), “Uses and abuses of patent statistics”, in A.F.J. van Raan (ed.), *Handbook of Quantitative Studies of Science and Technology*, Elsevier Science Publishers, Amsterdam.

Source URL: <https://www.innovationpolicyplatform.org/content/aspects-and-availability-patent-data>

Links

[1] <https://www.innovationpolicyplatform.org/content/patent-data-characteristics-innovative-activities?topic-filters=12070>

[2] <https://www.innovationpolicyplatform.org/content/patent-data-patent-valuation-methods?topic-filters=12279>