ARTICLE

Peter Evans and M. Khalid Nadeem Khan
Office for National Statistics

The characteristics of patenters

SUMMARY

This article describes how patents data from the Intellectual Property Office are being matched with detailed firm-level data from the Office for National Statistics. The resulting database will provide a gateway for future research on firm-level innovation and technical change. Preliminary results are presented showing the characteristics of patenting businesses, the rate of patent renewal, and the regional distribution of patenting activity. The article also highlights future areas of research that can develop the evidence base for patent and innovation policy.

Introduction

Patent data are an early indicator for understanding technical change and innovation. It has offered insight into scientific progress and the structural change of economies. The main focus of this article is to understand who patents and what is being patented.

The inter-linkage between patents, research and development, trade in intangible assets and various other issues, at the national and international levels, has drawn the attention of policy makers. Policy makers seek to improve the effectiveness of the patent system to promote innovation, and the policy interest motivates a deeper investigation of the patent data.

Such investigation requires detailed business level data, not only for patents and patenters, but for research and development, international trade in intangible assets and business structure. These data are not easy to collect and such a process is very costly. An indirect way of achieving the same objective is to use existing data from various sources and match it to patent data.

The Office for National Statistics (ONS) and the Intellectual Property Office (IPO) have started a joint project to create a dataset by matching patent data with ONS business data. The initial stages of this project have been completed and the results of the match are promising, indicating that the dataset can be used to answer many policy questions.

This article covers the methodology of matching, provides some preliminary

analysis, and indicates how important policy questions could be answered through matching with other ONS datasets.

The article is organised into the following sections:

- a description of the ONS and IPO data
- step-by-step details of the data matching
- a preliminary analysis of the matched data to investigate the characteristics of patenters: looking at patenting businesses, patenting intensity and the regional distribution of patents
- an analysis of the patent renewal rate, considering the life-length of patents
- a look at the potential for further matching and the usefulness of matched data
- concluding remarks

IPO and ONS data

Before describing the matching process, it is useful to understand the nature of the two datasets and their affect on matching. This section covers the IPO and ONS datasets, with regard to the coverage and information they provide.

IPO Data

The IPO is the legal body that grants patents within the UK, and therefore is the main source of patent information for the UK. For this study, the IPO provided data that contained 35,384 patents from January 1999 until September 2008.¹ This dataset only covers businesses and provides information on successful

patents. These are patents that were filed and subsequently granted.

The IPO data provide the following information:

- agent, applicant and proprietor name
- patent filing/application date²
- patent granting date, and
- patent renewal information

It is useful to explain a few of these terms. The agent is an individual or business that applies for a patent on behalf of the inventor. The applicant is the organisation on whose behalf an agent applies for the patent. The proprietor is the legal owner of the patent, which is either reported on the patent application or is the entity that has renewed the patent. The economic owner is of interest for policy purposes because it identifies the inventor and subsequently the beneficiary of returns from the patenting rights. Therefore, this discussion and preliminary results use proprietor as a point of reference, unless stated otherwise.

Since a UK patent currently takes an average of three years from filing to granting, it is useful to understand an exception to the normal patent process. In cases where one application is further divided into two patent applications (known as 'divisionals'), the second patent is given a new filing/application date, reflecting when this application was received, but the actual filing/application date remains the date when the original patent was filed. This complication needs to be treated and interpreted carefully. **Figure 1** presents the number of patents filed in each year – as mentioned previously, this dataset contains information on patents that were subsequently granted.

Figure 1 depicts two interesting trends. Firstly, until 2004 the trend in patent applications is somewhat similar. Secondly, there is a significant decline from 2005, which requires further exploration. The IPO annual review claims a 5 per cent decline in the filing and granting of patents,³ yet this is not a satisfactory justification as Figure 1 shows a sharper decline. The answer lies in the patenting process. Since we have data until 2008, it is likely that a sizeable number of patents are still in the application process and yet to be granted.

There is known to be a level of inaccuracy within Figure 1 as this dataset uses the filing/application date. As stated previously, this is accurate unless there is a preceding filing/application date. A preceding filing/application date is required for

divisionals or patents filed under the Patent Cooperation Treaty (PCT). When a patent is applied for through the PCT system, many steps of the patenting process are completed prior to the application reaching the national patent office. Therefore, the priority date for PCT applications is the date of the patent application made in the original country.

This inaccuracy has further implications for the analysis presented within this article. At present, it is not possible to determine how many patents are divisionals or were applied for via the PCT, which results in a possible under/over statement within each period⁴.

Figure 1 is a snapshot of annual patenting activity, but the level of analysable data in the dataset allows for a life-length perspective of patent stock. **Figure 2** shows the stock of patents in the matched dataset, depicting how many of the 35,384 patents are added to the patent stock and how many die within a given year. Patents are required to be renewed after their first five years of existence (starting from the filing/application date), the first occurrence of patent death is in 2003.⁵

ONS data

The Business Structure Database (BSD) holds business information on the vast

majority of businesses operating within the UK (in 2006 the BSD covered 3.3 million businesses). The BSD is an annual download from the Inter-departmental Business Register (IDBR) which is a live database of business information based on HM Revenue and Customs (HMRC), Dunn and Bradstreet⁶ business information, ONS surveys and Companies House data.

The BSD provides the information shown in **Table 1**⁷.

It is worth noting that, once patent and BSD data are matched, further matching with other ONS datasets (such as the Business Enterprise Research and Development dataset, BERD) becomes relatively straightforward as all ONS datasets contain a point of reference attained from the IDBR.

Matching IPO and ONS data

In the absence of a common business identifier in both the datasets it is difficult to match IPO and ONS data. However, since both these datasets contain business names, a name-based matching method is used. It should be noted that each patent may have three business names (representing an agent, applicant, and proprietor). Therefore all three names are matched with corresponding names within the IDBR and, through this link,

Figure 1
Number of patents

Source: Intellectual Property Office

Figure 2
Patent activity

Source: Intellectual Property Office

Table 1
Variables contained within the BSD

Definition	BSD variables
IDBR identifier code	Enterprise ref
Reference of parent organisation	Enterprise group ref (wow)
Postcode	Postcode
Foreign ownership marker	Immediate foreign ownership
Foreign ownership marker	Ultimate foreign ownership
Employment for enterprise	Employment
Turnover for enterprise	Turnover
SIC92, SIC03, SIC07 (depending on year)	Standard Industrial Classification
Year business began trading	Birth date
Year business stopped trading	Death date
Number of local units that make up the enterprise	Number of local units
Number of reporting units that make up the enterprise	Number of reporting units
Legal status of enterprise (e.g. company, partnership, sole proprietor)	Status

Source: Office for National Statistics

Table 2
Matching results of IPO and ONS data

IPO data category	Number of patents	Percentage of patents
Matched data		
Agent, Applicant and Proprietor	12,298	34.76
Applicant and Proprietor	729	2.06
Agent and Proprietor	1,161	3.28
Agent and Applicant	731	2.07
Proprietor only	243	0.69
Applicant only	55	0.16
Agent only	17,759	50.19
Subtotal	32,976	93
Unmatched data		
IPO data that could not be matched to business name	2,408	6.81
Total	35,384	100

Source: Intellectual Property Office and Office for National Statistics

are matched with an enterprise reference number. This section documents the matching process in detail:

Data matching

The IDBR team was provided with:

- the name and address of the entity that filed the agent
- the name of the applicant
- the name of the entity listed as proprietor
- IPO unique patent number (available for agents) or a unique reference number created for applicants and proprietors

Subsequently, the patent data was split into two subsets: one with agents, another with applicants and proprietors. In order to improve the probability of matching, the names of patenters were standardised. For example, *Company X Limited*, *Company X plc* and *Company X* would be altered to *Company X*. This reduces the number of observations that cannot be matched because of different name recording conventions, and minimises time as fewer names are matched. In the final step, the matched file, based on name, is reattached

to the original IPO dataset using the unique patent numbers mentioned above.

The patent dataset

After matching by name, the patent dataset has an IDBR enterprise reference number for agent, applicant and proprietor⁸. Each observation also contains information regarding:

- date of application of the patent
- date that the patent was granted, and
- the age of the patent if older than 5 years

The inclusion of the enterprise reference number allows the final piece of matching, between the patent dataset and the BSD.

Matching to the BSD provided further complications due to the non-uniqueness of businesses in both datasets. The patent dataset has multiple observations for businesses that own more than one patent. Similarly, the BSD has multiple observations for the same business if it exists in more than one period. Both situations are very likely and would cause considerable problems if matched directly together. In such a situation, it is important to simplify the matching process. In this

instance, the patent data were given a unique identifier (such as numbering the data from 1 to 35,000). Linking the enterprise references (and unique identifier) from the patent data to the business data (a relatively straightforward match) allowed a subsequent match; linking the business data to the patent data via the unique identifier. This process is straightforward and avoids the difficulty of sorting through multiple observations.

The IPO–BSD dataset

The matching of the two datasets was relatively successful; 93 per cent of IPO observations were matched (to some degree) with the BSD. The IPO data did not match for all points of reference: agent, applicant, and proprietor. **Table 2** presents the results in terms of matched and unmatched observations.

As stated previously, analysis should focus on proprietors as they are the beneficiaries of the patent, and because such analysis would be of use for policy purposes. Table 2 shows that about 40 per cent of proprietors matched to ONS data and are, therefore, useful for further analysis. The results indicate that, with possible improvements in the data from the IPO and investigation into bettering the name-matching algorithm, a greater success rate could be achieved. These results are very encouraging and indicate the possibility of further research to answer policy questions through matched data.

The characteristics of UK patenters

This section analyses the matched data to understand the characteristics of businesses patenting in the UK. Specifically, it tries to understand patenting activity with reference to business size. As stated above, the proprietor is the focus of analysis because this closely identifies the beneficiary of the patent.

This analysis does not exhaust the possibilities of the combined dataset, but demonstrates the detailed findings that can be produced. Generally, the relationship between business size and patenting activity attracts a lot of attention in policy circles.

Similarly, information on the location of patenting businesses is useful for regional economic and innovation policy. This article shows the potential of using the matched datasets to understand the location of patents and other related questions.

Size analysis

The innovation and patent literature highlights that patents are a good proxy for innovation (Helmers and Rogers, 2009). Therefore, to ask, what type of businesses patent to a greater degree? is tantamount to asking, does variation in business size explain variations in patenting habits (Balasubramanian and Sivadas, 2008)? Consequently, a preliminary analysis of patent activity across different sizes of business could potentially reveal interesting patterns of patenting behaviour.

In this section, business size is based on UK audit regulations. The regulations state that there are three measures on which to determine the size of a business. This analysis will look at two of these, employment and turnover⁹. First, the patenting businesses are presented according to their size. This means that, if a business patents, the analysis focuses on the size of businesses irrespective of the number of patents they own. Subsequently, focus shifts to patents, explaining how many patents are filed by businesses of different sizes. Effectively, this analysis shows the size and the degree to which businesses of a given size patent.

Size analysis – size determined by number of employees

Figure 3 shows that most of the businesses in the database (about 55 per cent), are small in size from an employment perspective.

Despite the limited period covered within the dataset, **Figure 3** shows that, over time, the proportion of small businesses that patent has increased and the proportion of large businesses has fallen. This is an interesting observation and further investigation to understand what has caused this change is needed. In addition, it would be useful to understand the resulting implications for patent and innovation policy.

Patent intensity is a more useful aspect to analyse because of its significance for patent and innovation policy. **Figure 4** depicts the patent intensity by business size analysis. Patent intensity analysis answers the question of ‘what type of business owns the most patents?’ Despite the presence of a high number of small businesses with patents, the lower number of large businesses account for, on average, a slightly greater proportion of total patenting. The most interesting aspect of this is that the proportion of patents held by medium-sized businesses is

comparatively low, on average only 16 per cent of patents.

With additional matching, further analysis could be undertaken to investigate the difference between patenting and non-patenting businesses.

Size analysis – size determined by level of turnover

The number of employees in a business is a good proxy for size but may not be relevant to all sectors, especially in businesses where the production structure favours small numbers of employees, or economies of scale require a business to be large. To control for this effect, similar analysis has been undertaken using turnover as an indicator of a business size.

Turnover could be considered a better proxy for financial strength and business performance. **Figure 5** presents the percentages of businesses in each size category. There is only a slight difference in the number of businesses that have patents when comparing size based on turnover with size based on number of employees. The relative proportions of small, medium and large businesses are

also similar, providing little extra insight. Therefore, it could be assumed that, businesses with fewer employees also have low turnover and qualify as small on both measures.

The patent intensity when turnover determines size differs from when size is determined by employment. **Figure 6** shows that large businesses have a higher patent intensity, on average 5 percentage points higher, when size is determined by turnover rather than employment. The intensity of patenting for small businesses is similar between the two determinants of size. Therefore, the proportion of patents held by medium-sized businesses is 5 percentage points lower when turnover, not employment, determines size.

The differences, though not major, show that the grouping of businesses for size analysis is sensitive to the criteria selected. Using more than one criterion is useful because it enriches conclusions and provides other analytical perspectives.

These results raise the question, what factors are responsible for the patenting intensity of small and large businesses? A complete explanation may include

Figure 3
Business size in the patent data

Source: Intellectual Property Office and Office for National Statistics

Figure 4
Patent activity by business size based on employment

Source: Intellectual Property Office and Office for National Statistics

Figure 5
Business size by turnover

Source: Intellectual Property Office and Office for National Statistics

Figure 6
Patent activity by business size based on turnover

Source: Intellectual Property Office and Office for National Statistics

other important factors such as business life, maturity, sector and the products of the business. Small businesses may be advantaged because of their flexibility and concentrated focus in a specific area. However, large businesses may have economies of scale that give them an advantage in creating and patenting. These suggestions are interesting, but further research is required to fully understand what determines the intensity of patenting.

Regional analysis

The regional dimension of patent activity is useful because it provides information for regional economic policy as well as innovation policy. Literature on innovation and patents argue that patents are a good predictor of innovation and therefore play a vital role in determining national and regional economic growth, company performance and the ability of businesses to compete internationally.¹⁰ To understand this dimension, we have used patent location from ONS data to observe patenting activity within the UK. Though the results are preliminary, they are indicative of regional trends in patenting.

The regions are classified according to a patenters' Nomenclature of Units for Territorial Statistics (NUTS) region. NUTS regions classify the UK territory based on government offices, groups of counties and groups of unitary authorities. Due to data limitations and disclosure restrictions, the analysis focuses on NUTS2 region (groups of counties).

Map 1 shows the average number of patents from 1999 to 2008, within NUTS2 regions. The map shows that the majority of patenting takes place in the South East of England. This could be due to the location of London. Interestingly there were also large numbers of patents in East Wales and Eastern Scotland which contain Cardiff and Edinburgh.

It is important to consider that this analysis reflects the location of the enterprise that patents, not necessarily the part of a business that innovates. This is due to the BSD containing the postcode for business, not necessarily the unit within the business that patents. For example if a business unit in Newcastle registers a patent but the enterprise who owns this unit is located in London, this analysis considers

the patent to have been generated in London. Hence, there could be a head office bias that may be particularly prominent when considering capital cities.

The cluster of Berkshire, Buckinghamshire and Oxford is an interesting region because, at the NUTS2 level, it has (on average) the highest number of patents. This could be explained partially through its close proximity to London, business linkages and Oxford University. Likewise, the presence of Cambridge University might account for the considerable patenting activity in East Anglia. This preliminary analysis shows that further investigation will be useful to provide understanding for patenting trends and could contribute to evidence for policy formulation.

Patent renewal rate

The rate of patent renewal is of particular interest to patent organisations around the world and has been a major point of discussion in literature (Van Zeebroeck, 2007, Pakes and Schankerman, 1984). It provides evidence on the life-length of innovation and the impact of renewal fee on the patent renewal rates. Furthermore, patent renewal rate is used to estimate research and development (R&D) and other intangible assets' life-lengths (Wenzel et al 2009). This section offers preliminary results on patent renewal rate at aggregate level.

The data covers nine years, from 2000–2008, therefore there are only four years after the 5 years of initial protection expires, that could provide evidence on the rate of patent renewal. This imposes a limitation on the results but still offers some insights.

It is important to point out that this section is particularly affected by patents such as divisionals and those applied for under the Patent Cooperation Treaty (PCT) that, within this dataset, have the filing/application date of the second patent/UK application date, rather than the filing/application date of the first patent/foreign application date. This affects the results as the renewal rate of patents may be under or overstated. While this hampers the usefulness and dependability of results, it does provide some indication regarding trends and the ability to undertake such analysis in future work.

Figure 7 shows the patent renewal rate in the total patent stock. This analysis treats each patent as a data point and then traces its life trajectory. After year five, when the patent renewal fee becomes due,

Map 1

Average number of patents by UK NUTS 2 area, 1999 to 2008

Note:

¹ The counts for Highlands and Islands, North Eastern Scotland and Cumbria are below the disclosure threshold of 10 businesses and are excluded from the map.

Source: Intellectual Property Office and Office for National Statistics

Figure 7
Renewal rate

Source: Intellectual Property Office and Office for National Statistics

the percentage of patents renewed steadily decreases. There are no dramatic changes in the renewal rate which falls steadily over time. Obviously, a longer time series would provide a better picture of the renewal pattern, but analysis across industries and businesses could also contribute to the understanding of variations in renewal rate. Subsequent work will focus on such cross-section analysis.

Matching to other business data registers

This article provides evidence that analysis on a patent/business dataset is practical and can provide insightful information. Such a dataset provides the opportunity for further matching to other ONS datasets. The resulting data could address specific policy questions on R&D and patenting, innovation and patents, and international trade of intangible assets and patents, among others.

Investigating the characteristics of businesses that patent is an important area for development, in order to understand why some businesses are more successful at patenting than others. This section provides a brief overview of some of the datasets that are being considered as part of a future plan for matching and analysis.

The Business Enterprise Research and Development (BERD) dataset covers expenditure on R&D activities. A match between IPO and BERD data could help to answer questions on links between R&D and patenting. R&D is likely to be the source of the vast majority of a business's patents. Therefore it would be of interest to analyse the effects of changes in R&D expenditure (for example, volume) upon the number of patents produced. This would not only examine the volume of patenting generated from R&D, but also provide evidence for the success rate of R&D and/or the perceived usefulness of innovative businesses patenting.

The International Transactions in Intangible Assets (ITIS) dataset provides information on the sale of various intangible assets, many of which would be protected by patents (particularly intangible assets such as R&D and computer services, among others). Analysis of this dataset could provide information on the value of patents to a business in terms of the sales generated of patentable products/services. Hence, this could provide a method of measuring the value of patents, an area of great interest to patent offices and academics alike.

The Annual Respondents Database (ARD) provides a comprehensive breakdown of information on business activity: expenditure on inputs, human resources, sales and so on. Matching this to IPO data can provide insight into the level of turnover or expenditure of patenters. This information could provide understanding of what differentiates patenting and non-patenting businesses, and what determines the intensity of patenting.

The Community Innovation Survey (CIS) focuses upon innovation within businesses. This survey includes questions on the importance placed on activities such as patenting and R&D, therefore providing primary evidence on the link between patenting and innovations. The CIS can provide information on patenting premiums, through the profitability of patenters compared to other innovative businesses, and how innovation is protected: through patenting or other strategic methods like secrecy contracts.

The analysis in this article has used business information obtained from the BSD. This dataset is not a spent resource. Perhaps the most notable piece of work that could be undertaken is an investigation into how the level of patenting is affected by the age of a business. Such an investigation is currently being undertaken in France where, among other things, it was

discovered that patenting propensities depend on the age of a business (Lelarge, 2009; this paper is in preliminary stages and has yet to be published).

The use of such datasets, either individually or collectively, would provide a rich insight furthering current research into patenting and patenters. Various academics and international bodies are undertaking such research.

Conclusion

This article indicates that matching patent and business data can produce useful analysis that could meet the needs of patent and innovation policy makers. The current analysis is limited and considered preliminary; however, further investigation into data matching and types of analysis should be undertaken to improve the current dataset.

Developing matched databases is in the interest of patent organisations, national statistics institutes and policy makers in general. Patent organisations can gain a deeper knowledge of their customers' needs, which can feed into the patenting policy. This dataset allows access to a comprehensive patent-business dataset with the potential to be matched to many other ONS datasets, able to provide analysis on specific points of interests.

Comparing this dataset to currently created databases that match company data with patent data, offers an opportunity to compare businesses and companies and, if similar, the ability to use investigations on one particular dataset as a proxy for the other. Analysis of companies benefits from having a vast amount of company information, but does not consider all businesses. Analysis of businesses will consider all companies and businesses, and uses different business information.

The preliminary analysis in this article identifies that the majority of patent activity is undertaken by a small number of large-sized businesses and a large number of small-sized businesses. Furthermore, it would seem that UK patent activity occurs predominantly in the South East of England. The renewal rate of patents is consistent throughout the period for which information is held.

To conclude, this article clearly identifies a methodology for matching patent and business data. Such a match can be undertaken with the vast majority of observations matching in some form. Furthermore, such a dataset allows a detailed and varied analysis of patenters and patenting for UK-based businesses.

Notes

1. The dataset contained a few observations from 1998 but due to very low coverage of the 1998 patents we have dropped this year from matching.
2. Filing and application means essentially the same thing and we use these terms interchangeably. Filing is a preferred term in the IPO while the OECD Patent Statistics Manual uses application to describe the same process.
3. www.ipo.gov.uk/about/whatwedo/ourpublications/ourpublications-review.htm
4. The regional analysis used an average number of observations across the period; therefore the use of the incorrect application year should have little or no affect.
5. A handful of observations relating to patents applied for in 1998 were present within the dataset. The year 1998 was not included in analysis, but these patents were. Hence, this is why the first patent death occurs in 2003.
6. www.dnb.co.uk/about-dnb.asp
7. For more details on BSD see Evans and Welpton (2009).
8. If a particular patent had a point of reference, for example the proprietor, that was not a business or could not be matched to an enterprise reference number by the IDBR team, then this part of the patent data could not be matched to the BSD
9. Business size according to audit regulations:

	Small	Medium	Large
Employees	less than 50	51 to 250	Greater than 250
Turnover	£5.6 m	£5.61m to £22.8 m	Greater than £22.8m

The other measure of business size is gross assets, which is not attainable from this dataset.

10. Rosenberg (2004) specifies that technological innovation is a major force in economic growth.

CONTACT

✉ elmr@ons.gov.uk

REFERENCES

- Balasubramanian, N. Sivadasan, J. (2008) 'What happens when businesses patent? New evidence from U.S. economic census data'. Center for Economic Studies, discussion paper January 2008.
- Evans, P., Welpton, R. (2009) 'Methods explained: Business Structure Database' *Economic and Labour Market Review*, vol 3, no 6, pp 71–75.
- Evans, P. Hatcher, M. Whittard, D (2008) 'The preliminary R&D satellite account for the UK: a sensitivity analysis' *Economic and Labour Market Review* 2 (9) pp37–43
- Galido-Rueda, F. (2007) 'Developing a UK satellite account for UK: A preliminary analysis' *Economic and Labour Market Review* 1 (12) pp18–29
- Helmets, C., Rogers, M. (2008) *Innovation and the Survival of New Firms Across British Regions*. University of Oxford, Department of Economics in its series *Economics Series Working Papers*, number 416
- IPO (2009) 'Facts and Figures' at: www.ipo.gov.uk/about/whatwedo/ourpublications/ourpublications-review.htm
- Lelarge, C (2009) 'The innovative activity of firms over their life cycle: Evidence from French mirco-data'. Paper presented at the Comparative Analysis of Enterprise Data (CAED) conference Tokyo, Japan, 3 October 2009, available at: <http://gcoe.ier.hit-u.ac.jp/CAED/program.html>, accessed 12 November 2009
- Pakes, A. Schankerman, M (1984) 'The rules of obsolescence of patents, research gestation lags, and the private rate of return to research resources' in Zvi Griliches (ed.) *R&D, Patents and Productivity*, Chicago, University of Chicago Press, pp 73–88. British Library shelf mark 84/21211
- Rosenberg, N (2004) 'Innovation and Economic Growth' at: www.oecd.org/dataoecd/55/49/34267902.pdf
- Van Zeebroeck, N. (2007) 'Patents only live twice: a patent survival analysis in Europe' CEB Working Paper No 07/028 October 2007.
- Wenzel, L. Khan, M.K.N. Evans, P (2009). 'Capitalising research and development: towards the new system of national accounts'. *Economic and Labour Market Review*, vol 3, no 9, pp 16–23.