Jointly published by Kluwer Academic Publishers, Dordrecht and Akadémiai Kiadó, Budapest

Scientometrics, Vol. 49, No. 3 (2000) 511–515

Short communication

A note on using patent statistics to obtain competition indicators

SHYAMA V. RAMANI, MARIE-ANGELE DE LOOZE

INRA, Grenoble Unit, Université Pierre Mendès France, Grenoble (France)

In a recent article a set of indicators have been proposed drawing upon patent statistics, which are meant to describe and compare firm and national research competence. However this article has raised more questions on the validity of such indicators as well as on their use. We have thus examined these issues so as to clarify the nature of the problems involved in the construction of competence and competitive indicators of firms and nations and their subsequent implementation on data bases.

In a very recent article in *Scientometrics*, *Banerjee* et al. have proposed a set of indicators drawing upon patent statistics, which are meant to describe and compare firm and national research competence. They have illustrated their concepts with an analysis of patent data on the biotechnology sectors. However there seem to be some problems both with the interpretation of these indicators as well as their application to the specific data base concerned. Thus the objective of this short note is to discuss these issues so as to initiate debate on how patent statistics may be used to reveal the degree of firm or national research competence.

We begin by first reformulating the indicators proposed by *Banerjee* et al. in a general context. Suppose in a data base m types of patent observations can be distinguished. Furthermore, suppose that the observations are available for n number of years. Then consider the following terms that can be derived from such a data base.

Let t^i_j represent the total number of patents of type i in year j. Let T_j be the total number of patent observations in year j. Let T^i be the total number of patent observations of type i over the n years. Let T be the total number of patent observations in all (i.e. of all types over all years).

0138–9130/2000/US $ 15.00

Note that t^i_j and T_j are dependent on the year considered and vary from year to year. On the other hand T^i and T are constant over time as they are the averages over the time period considered. Then the following indicator I^i_j can be constructed for each type i and for each year j:

$$I^i_j = \frac{t^i_j / T_j}{T^i / T} = \frac{\text{Share of patent type i in year j}}{\text{Share of patent type i over all years}}$$

The numerator shows, the share of the patent observations of type i, for each year. It is a positive number between 0 and 1. The denominator shows the share of type i patent for the entire period considered and this is also a positive fraction. The numerator is time dependent as it varies for every year. The share of patent type i in year j indicates the weight of type i in the year j. The denominator is time independent as it is a constant. The final indicator I^i_j can be greater than or less than 1, as it is the ratio of two fractions. What can be deduced about competence or competitive position of type i from such an indicator? We can infer that in year j if the indicator I^i_j is greater than 1, it will decrease in the future and vice versa, because when the share of patent type i in year j is greater (or less) than the share of patent type i over all years, then type i patent has to decrease (or increase). Apart from such inferences it is not clear how such indicators can be used to measure the competitive position of type i vis-a-vis other types.

The typologies that have been examined in *Banerjee* et al. are languages, sectors and countries. Sometimes "half-lives" are mentioned. It is not clear how this concept can be applied to a variable that is not a well defined and well identifiable function of time.

Variants of the above have been developed for countries and these are meant to represent their competitive position, dependency and self-sufficiency. In terms of the above notation, consider i to represent country i. Then the following indicators are proposed:

$$\text{Competitive position of country i in year j} = \frac{t^i_j \Big/ (T_j - t^i_j)}{(T - T^i)\Big/T} = \frac{t^i_j \Big/ (T_j - T^i_j)}{(T_j - t^i_j)\Big/T}$$

$$= \frac{\text{ratio of country i's production in year j to total production of rest of the world}}{\text{ratio of production of rest of world in year j to total production of the world}} .$$

$$\text{Auto-sufficiency of country i} = \frac{T^i}{T} \quad .$$

$$\text{Dependency ratio of country i} = \frac{1 - \left(T^i \big/ T \right)}{T_i \big/ T} = \frac{T - T^i}{T^i} \quad .$$

Now the above measures are based implicitly on a game with only two players; the country i itself and the rest of the world. It is doubtful whether countries formulate strategies or targets with respect to the rest of the world. It seems more likely that they compare themselves with a set of their competitors. It is not apparent in what measure the given index can be justified as an indicator of the competitive position of a country. This needs to be related to economic theories of competition and specialization. Furthermore, as the authors themselves point out, it is difficult to justify and interpret the "auto-sufficiency" and "dependency ratio" given above. Perhaps, it would be simpler to understand them as "export" and "import" ratios of knowledge. For instance the "auto-sufficiency" can be taken to measure the potential export of knowledge, if patents can be accessed and if they serve to diffuse knowledge rather than prevent the circulation of knowledge. Following the same line of reasoning, the "dependency ratio" would then simply be the import to export ratio of knowledge of country i. It is extremely imprudent to attach notions of dependency or sufficiency to such measures without justification. It is interesting such measures have also been used to indicate "spillovers" or "free knowledge flows" from developed countries to developing countries, which seems equally contestable.[2]

We now look at the applications of the above indicators ("language pool", "sectors" and "country positions") to the data base on biotechnology. The data base concerned is the "Derwent Biotechnology Abstracts" and it has been incorrectly spelt throughout.

The language in which a patent is applied for often depends on the region of protection sought by the patentee. For instance, if a Japanese company aims for national protection it will patent it in the Japanese language. However if it wants world wide protection, it needs to patent it in English. Therefore the ranking of language pools itself does not reflect the ranking of these languages in the patent observations. The reasons for the international dominance of English are obvious, but the ranking of the other languages is bound to differ, if world wide patents are distinguished from others. Finally, the use of ranking of language pools is not clear as it cannot credibly throw light on either national or firm research competence.

The "sectoral positioning" as derived in *Banerjee* et al. has to be also reconsidered. For instance the authors have compared knowledge production in the pharmaceutical and agricultural sector by comparing the patent observations in these two industrial fields. This brings up the question: should the production of patents in a sector, say the pharmaceutical sector, be gauged from the patent applications classified under the pharmaceutical sector or should it be measured by the patent applications of actors (firms and labs) active in this sector? It must be noted that there are two types of technology: "generic or pre-competitive" and "specific technology". Now the mass of patents in the biotechnology sectors is in genetic engineering and related fields. This has very strong links and implications for the future of the pharmaceutical sector because many of these depositions classified under genetic engineering are in fact by pharmaceutical firms. Therefore it is difficult to make pronouncements on the pharmaceutical sector without taking into account the range of patents deposited by the pharmaceutical companies. Lastly, Derwent changed its classification scheme in 1992 and it is not clear how such changes have been incorporated in the data base.

The "country competition" results are also difficult to accept without questioning because the "Derwent Biotechnology Abstracts" does not indicate the national affiliation of the patentee or the group of patentees who submit an application together. Therefore the national affiliation depends on whether it is inferred from the "patent number", the "priority number", the "location of patent assignee" or by some other means. The patent number indicates the region of final protection sought, while the priority number indicates the initial region of protection sought. For instance if Institut Pasteur which is located in France, applied for regional protection (EU) and then for world protection, in the patent records it would appear associated with three kinds of regional affiliations. For many observations the national affiliation is simply not clear and cannot be deduced. There are also other serious theoretical questions. For example, if a subsidiary of Unilever in India, say, Hindustan Lever takes out a patent in the USA is this a reflection of the research capacity of India or of the Netherlands?

Finally, many of the indicators are indexed to time or to a particular year. This index should also be treated with caution. To start out, a distinction has to be made between patents that have been "granted" and patents that have been "applied for". The data base gives the date of application and the date of patent publication. The "priority number" indicates the initial date of application while the "publication number" indicates the date when the patent was published. The publication date is not a credible indicator of time because of heterogeneous international practises. In the case of the USA the date of publication is the same as the date of patent granted, because in the USA patents are published only if they are granted. In Europe, patents are systematically published

18 months after the application and the data base does not indicate whether they have been granted in all cases. There is a large variance in the number of years between patent application and patent publication. Therefore the "priority number" is the only credible indicator of time but again it refers to patents applied for but not necessarily granted.

In conclusion, we reiterate that *Banerjee* et al. have raised an important question: how can we use patent statistics to measure firm or national research competence? To this end they have proposed a set of indicators and applied to them to the "Derwent Biotechnology Abstracts" data base. However their article has raised more questions on the validity of such indicators as well as on their use. We have thus raised these issues so as to clarify the nature of the problems involved in the construction of competence and competitive indicators and their subsequent implementation on data bases, in order to stimulate further reflection and work on these questions.

References

1. BANERJEE P., B. M. GUPTA, K. C. GARG, Patent statistics as indicators of competition an analysis of patenting in biotechnology, *Scientometrics,* 47 (2000), 95–116.
2. BASANT, R., B. FIKKERT, The effects of R&D, foreign technology purchase, and domestic and international spillovers on productivity in Indian firms, *Review of Economics and Statistics*, (1996), 187–199. DEOLALIKAR, A. B., R. E., EVENSON, Technology production and technology purchase in Indian industry: An econometric analysis, *Review of Economics and Statistics*, (1989) 687–692.

Received August 8, 2000.

Address for correspondence:
SHYAMA V. RAMANI
INRA, Grenoble Unit, Université Pierre Mendès France,
B.P. 47, 38070 Grenoble cedex 9 (France)
E-mail: ramani@grenoble.inra.fr