Technological Forecasting & Social Change 81 (2014) 287–299

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

Selection bias in innovation studies: A simple test

Gaétan de Rassenfosse [a,*], Anja Schoen [b], Annelies Wastyn [c]

[a] The University of Melbourne, Melbourne Institute of Applied Economic and Social Research, and Intellectual Property Research Institute of Australia, Level 5, FBE Building, Victoria 3010, Australia
[b] Technische Universität München, TUM School of Management, Arcisstr. 21, 80333 Munich, Germany
[c] KU Leuven, Dept. of Managerial Economics, Strategy and Innovation, Naamsestraat 69, 3000 Leuven, Belgium

ARTICLE INFO

Article history:
Received 9 July 2012
Received in revised form 18 December 2012
Accepted 27 February 2013
Available online 22 March 2013

JEL classification:
O31
C18
C52
C81

Keywords:
Knowledge production function
Patent count
R&D
Selection bias

ABSTRACT

The study of the innovative output of organizations often relies on a count of patents filed at one single office of reference such as the European Patent Office (EPO). Yet, not all organizations file their patents at the EPO, raising the specter of a selection bias. Using novel datasets of the whole population of patents by Belgian firms and German universities, we show that the single-office count results in a selection bias that affects econometric estimates of invention production functions. We propose an easy-to-implement methodology to evaluate whether estimates that rely on the single-office count are affected by a selection bias.

1. Introduction

Economic and management research on innovation has greatly benefited from the increased availability of patent data, which provide a unique way of tracking the creation and the diffusion of innovation. Patent data are now being used in other fields including labor economics, development economics and economic geography. Yet, the measurement of innovation using patent data suffers from limitations. The two most severe of these limitations are that: (i) not all inventions are patentable and not all patentable inventions are patented; and (ii) the value of patents varies widely and the majority of patents are worthless. We refer the reader to [1–4] for in-depth discussions of these issues.[1]

This paper focuses on a third limitation, which is the selection bias that arises from the way patents are counted. It is common practice to count patents at a single patent office to assess organizations' inventive output (henceforth referred to as the 'single-office count'). A close look at a random sample of 20 scientific articles that use patent data, which were published recently in general economic and management journals as well as field journals, reveals that the overwhelming majority of studies rely on a single office count. However, this practice may result in selection bias

* Corresponding author. Tel.: +61 3 9035 4897; fax: +61 3 8344 2111.
E-mail addresses: gaetand@unimelb.edu.au (G. de Rassenfosse), schoen@wi.tum.de (A. Schoen), annelies.wastyn@econ.kuleuven.be (A. Wastyn).

0040-1625/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.techfore.2013.02.012

[1] Researchers working on the economics of innovation are well aware of these limitations. A typical solution involves using additional innovation indicators such as the number of scientific publications or the share of sales from newly-introduced products. Alternative indicators are not always available and many studies rely exclusively on patent data.

since firms have the option of filing patents anywhere in the world. This is particularly true in Europe, where two overlapping patent offices coexist. Companies may file patents directly at their national patent office or they may take the more expensive 'European route' by filing patents at the European Patent Office (EPO). Companies that target an international market may file their patents at the World Intellectual Property Office (WIPO) in Geneva, at the US Patent and Trademark Office (USPTO), or in any other jurisdiction. As long as filing decisions are random, the single-office count is a noisy proxy of the full patent count (i.e. the count that encompasses patents from all possible patent offices). However, as soon as systematic factors affect decisions to select a given filing route, the single-office count results in a selection bias.

Motivated by the tension between the popularity of the single office count and the threat of a selection bias, this paper proposes a way to test the existence of bias when the researcher observes patents at only one patent office. Using novel data on Belgian patenting firms and German universities, we show that the single-office count biases econometric estimates of invention production functions. Invention production functions relate an organization's inventive input to its output and are a key object of analysis in the innovation literature. We also show that our test, which uses information that is readily available to most researchers, successfully spots variables that are subject to a selection bias. It should be of interest to a wide audience given its ease of use and the popularity of the single-office count.

The paper is structured as follows. The next section surveys current practices in the way to count patents to estimate invention production functions. Section 3 explains the proposed methodology to detect a selection bias and Section 4 presents the econometric framework. The test is put into practice using data on Belgian firms in Section 5 and using data on German universities in Section 6. Implications regarding data collection and the estimation methodology are presented in Section 7, together with concluding remarks.

2. Measuring inventions with patent data: from theory to practice

Patent data are used in various ways and the appropriate patent indicator necessarily depends on the research objective. Here, our focus is on building a patent indicator to estimate invention production functions, a popular object of analysis in the innovation literature. Invention – or knowledge or patent – production functions relate organizations' research inputs such as R&D expenditures to their patented output. They have attracted considerable attention in the literature, dating back to [5], and have been used, among other things, to study the occurrence of innovation (e.g. [6–8]); to study the invention process and the effectiveness of innovation policies (e.g. [9–11]); or as an intermediate step to study the determinants of productivity (e.g. [12,13]).

A patent provides protection only in the country in which it is filed. As a result, firms that want to protect their invention in different countries must file a patent in each relevant national patent office. The first patent describing the invention is called the 'priority filing', while the subsequent patents extending the protection in other jurisdictions are called 'second filings'. We use the terms 'priority filing' and 'priority patent application' interchangeably. The priority patent application is usually filed at the home patent office, although it could be filed at another patent office (the most popular being the USPTO, the EPO and the WIPO). Because applicants have a variety of patenting routes available to them, the patent count should theoretically include all *priority* patent applications filed anywhere in the world, regardless of the patent office of application. This global count of priority filings is explained in great detail in [14].

In practice, however, the operationalization frequently departs from this ideal situation. In particular, the count of patents is usually limited to a count at one reference office, usually the national patent office or the EPO for European firms. We studied a random sample of 20 papers that estimate patent production functions on European data and that were published in the recent past in general economic and management journals as well as in field journals (see Table A in Appendix A).[2] We find that 75% of the papers rely on the single office count, and the EPO is taken as the reference office in most of these instances. Surprisingly, very little information on the patent indicators is usually provided. In particular, the priority status of the patent documents (priority filings or second filings) is discussed in only two cases. Limiting the count to patents filed at one reference office is a simple and convenient way to count patents. It is, however, necessarily prone to measurement errors since only a fraction of the total patented output is observed. This measurement error is a random error if it results in an estimate of effect which is equally likely to be above or below the true value, and the single-office count is simply a noisy measure of the true count. However, non-randomness in the measurement error would lead to a selection that biases the estimates of the patent production function.

The question of whether the single office count results in a selection bias has not been studied explicitly, although some authors have reported evidence that systematic factors affect the decision of filing route. Seip [15] provides statistical evidence for Dutch patenting companies. He reports that 80% of the Dutch companies that filed patents at the EPO or the WIPO in 2003–2007 were large companies (more than 200 employees). Yet, out of the 5000 Dutch patent-filing companies, only 6% have more than 200 employees, suggesting a large selection bias in terms of firm size: large companies are more likely than SMEs to file their patents at the EPO or the WIPO. de Rassenfosse and van Pottelsberghe [16] show that the driving force of national and international patents differs. While nationally-filed patents are more reflective of the propensity to patent, international patents such as EPO patents are more reflective of the productivity of research (see also [17]). At the patent level, anecdotal evidence of a potential selection bias is provided in [18]. Using a large sample of patents granted by the EPO between 1990 and 1995, the authors find that firms adapt their filing strategies according to the expected value of the patent. Jensen et al. [19] come to a similar conclusion using

[2] To build the sample we searched Google Scholar for journal articles that contain the keywords 'patent', 'production function', and the name of one European country (e.g. 'France'). We only kept articles that were published in A* or A journals according to the 2010 Excellence in Research for Australia (ERA) Ranked Journal List. Three exceptions are [42,46,54], which are B journals.