Research Policy xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Research Policy

journal homepage: www.elsevier.com/locate/respol

A novel technology-industry concordance table based on linked inventor-establishment data

Matthias Dorner[a,b], Dietmar Harhoff[b,c,d,*]

[a] Max Planck Institute for Innovation and Competition, Munich, Germany
[b] Institute for Employment Research (IAB), Nuremberg, Germany
[c] Ludwig-Maximilians Universität (LMU), Munich, Germany
[d] Centre for Economic Policy Research (CEPR), London, United Kingdom

ARTICLE INFO

Keywords:
Patents
International patent classification
Industry classification
Concordance
Linked inventor-establishment data

JEL classification:
L13
L20
O34

ABSTRACT

Mapping technologies into industries is frequently required in empirical innovation studies, but many concordances only provide coarse mappings. We develop a novel concordance between industries and technologies making use of linked inventor-employee data for a large sample of German patents filed at the European Patent Office between 1999 and 2011. Inventors listed on these patents are matched and disambiguated with German social security records. Employment data recorded in this database include detailed industry codes describing the industrial activities of the inventors' establishments. These linked microdata allow us to identify the precise industrial origin of inventions, combine them with technology classifications from the inventors' patents and to generate novel concordance tables. We evaluate our approach by comparing the concordances with existing work. Moreover, in an empirical application, we find that patent based indicators represent valid proxies for the innovation performance of industries and thus help to measure innovation in the absence of other survey based indicators. Service sector innovation activities, however, appear to be barely represented by the patenting activities of inventors and establishments.

1. Introduction

Empirical analyses of economic growth, industrial organization, productivity, trade and innovation often employ patent data to measure inventive activity in order to approximate different levels of technological use and technological change.[1] Industry level data on patenting is also informative for policy makers because knowledge output of industries can be used as an indicator to evaluate technology or industrial policies, which are usually designed along the lines of sectoral innovation systems or industrial value chains. In studies at the industry level, patents have to be matched to sectoral classifications of industries. Prominent patent classifications (such as the International Patent Classification, henceforth: IPC), however, are usually based on technological characteristics. While these technology classifications serve important purposes in the patent system, e.g. to support prior art search, they cannot be connected directly to industry classifications and industry data. As we show below, patents in a particular technology area may originate in a broad range of industries. Vice versa, inventors who are employed in establishments of a given industry may file patents in many different technological fields. High quality micro data capturing these empirical relationships between the industrial origins and patented technologies are thus a valuable and necessary resource for building informative concordance tables.

Several proposals for concordance tables which allow a mapping between industry and patent classifications already exist. Despite being helpful and heavily used tools for empirical research, two important issues arise in these works: First, many concordances are based on very specific and often small data samples that limit their external validity and applicability across countries, industry classifications and time. Second, in most cases, the data for producing such concordances are based on firm-level information that are matched with assignee information documented on patents (e.g. EPO and OHIM 2013). A major disadvantage of a firm-based concordance system arises from the multi-product nature and related organization of modern production. Large firms are active in multiple industries and markets. In firm data, however, industrial activity is determined primarily from global value

* Corresponding author at: Max Planck Institute for Innovation and Competition, Munich, Germany.
E-mail address: dietmar.harhoff@ip.mpg.de (D. Harhoff).

[1] See Griliches (1990) for a comprehensive discussion of the advantages and weaknesses of patent data as an economic indicator. A study of economic growth using patent data is Aghion et al. (2014, 2015). An example of a study of technology transfer building on patent data is Eaton and Kortum (2002). Glitz and Meyersson (2017) estimate industry level patent counts for a study of productivity differentials between West Germany and the GDR. Cross-country studies of the R&D patent relationship are Meliciani (2000) and Danguy et al. (2014).

https://doi.org/10.1016/j.respol.2018.02.005
Received 26 February 2017; Received in revised form 6 February 2018; Accepted 7 February 2018

added (or turnover) based on the firms' most important line(s) of business. Since these large organizations hold the majority of patents, the precision of concordances will typically be less than satisfactory if firm-patent linkages are used. Our approach avoids these problems. Another issue is that existing concordances often provide only an industry of use (IOU) interpretation about the relationship between specific products of industries and patented technologies. From a theoretical perspective, however, in many empirical applications such as studies of knowledge production, researchers might rather prefer precise industry of origin (IOO) characterizations, which relate to industry specific sets of knowledge and technological opportunities.

To allow for a more precise and comprehensive linking of technology and industries, we describe a novel approach based on linked inventor-establishment data for Germany. These data record the industrial activity of 148,793 establishments which employed the inventors at the time when their 235,933 were filed (priority filing dates). These unique data were generated by matching inventors listed on patents that had been filed by at least one German inventor at the European Patent Office (EPO) between 1999 and 2011 with administrative labor market data of the Institute for Employment Research (IAB). Our data represent more than 80 percent of the patent filings in the largest European economy for a period of more than one decade. The linked employer-employee database of the IAB records the industry classification of the employing establishment (rather than the firm) of inventors in the NACE system up to the precision of the five-digit level. Industrial activities of these units are determined exclusively from labor input data, i.e. the actual production tasks, research or service activities carried out in the specific unit.[2] Thus, compared with (linked) firm level data, these industry codes of establishments give us very accurate register based industry of origin information for each patent, which we use to generate novel concordance tables with technology information available at the same time from patents.[3]

We provide several plausibility checks of our empirical approach. First, tests show that relationships described by our concordances, especially the industrial origins of most technologies, remain quite stable over the decade we study. Second, restricting the data to a subset of completely matched inventor teams in order to account for a potential matching bias between larger and smaller teams, does not change the conclusions derived from full concordance. Third, we show that concordance tables derived from our linked inventor-establishment data differ from the ones typically used, in particular with respect to the details of the technology-industry relationships being captured. In the cores of the concordances, however, we also detect plausible similarities with our concordance and existing approaches. Finally, we argue that our approach towards constructing concordance tables is better suited to regular updating than most of the earlier work based on idiosyncratic samples. Both data sources used for the mapping are generated by administrative processes and thus are subject to continuous updates.

A comprehensive set of concordance tables provided as supplementary data will hopefully enable other researchers to enrich their empirical analyses with industry or technology data and support the creation of novel statistical indicators for policy analysis. With respect to the latter, we show that patent intensities by industries estimated using our concordance table are highly correlated with commonly used innovation indicators derived from survey and administrative data. This holds in particular for the manufacturing sector. If these alternative data sources are unavailable, patent based indicators, which are less costly to obtain, may provide a reasonable substitute. Service sector innovation, however, appears to be barely represented by statistical indicators estimates from patent data.

The remainder of the paper proceeds in six sections. Prior research on technology-industry concordances is surveyed in section two. In section three we provide details about our data and the methodology used for generating the concordance tables. Section four describes exemplary concordance tables. In section five we present a set of tests and empirical analyses of the concordance tables. Section six concludes.

2. Literature review

The analysis of innovation at the industry level often requires information as to which technologies are being employed. Since the 1980s, a number of proposals have been made to link industries, respectively industry classifications such as NACE (Nomenclature statistique des activités économiques dans la Communauté européenne) or ISIC (International Standard Industrial Classification of Economic Activities), to technology categories. Since technology classifications cannot be converted directly into industry codes, patent data have been used to construct technology-industry concordance tables.

One of the first attempts to link industrial sectors to particular technologies was made by Kronz and Grevink (1980). The authors intuitively classified the patent applications of five countries (DE, FR, GB, LU, NED) according to the NACE classification and provided a concordance based on these results.

A more structured attempt dates back to the work of the U.S. Patent and Trademark Office (USPTO) in the 1980s. The USPTO assigned patent sub-classes of the U.S. Patent classification (USPC) to 41 industry classes (U.S. Standard Industrial Classification). Since the assignments of patents to industries are in many cases ambiguous, the concordance used fractional assignments. A major limitation of the concordance is that it is only applicable to U.S. patents.

Everson and Putnam (1988) used unique data from the Canadian Patent Office to build another concordance matrix. Between 1972 and 1995, Canadian patent examiners had assigned patent filings to industry of origin and industry of use codes. Based on Canadian patent filings in the years from 1978 to 1984, a direct concordance linking IPC codes to IOO and IOU information was created. The resulting Yale Technology Concordance (YTC) links eight IPC sections with 25 industries (Everson and Putnam, 1988, Englander et al., 1988). Kortum and Putnam (1997) used the YTC to predict patent counts by industry for the years 1983–1993. Results revealed that the predictions are fairly reliable for the early years and also for a subset of U.S. inventors. However, prediction errors are relatively large for non-U.S. inventors and for patent filings published after 1998. The authors concluded that the relationship between technology fields and industries has changed over time and that the applicability of the concordance varies between countries.

Based on 280 German patent filings, Greif and Potkowik (1990) provided a concordance matrix for IPC classes and branches of trade (Wirtschaftszweige), a German national statistical classification scheme of industries. Especially the small sample size raises some doubts about the validity of the concordance in empirical applications. Moreover, the results cannot easily be translated into international industry classifications, such as NACE or ISIC.

Verspagen et al. (1994) have advocated the MERIT concordance table, which provides a link between IPC classes and the ISIC classification (22 manufacturing sectors) based on Finnish ISIC codes prepared

[2] Regulations of the Federal Employment Agency (FEA) require establishments of at least one employee subject to social security contributions to apply for a unique establishment identifier. Obligatory (annual) reports of employees to the social security administration must document this establishment identifier and include an up-to-date declaration of the economic activity of the establishment according to the effective NACE classification. Assignment of NACE industry codes must consider the economic purpose of the establishment and in particular the activities the majority of employees perform. Criteria used by statistical offices to determine industries at the firm level such as, e.g., value added or turnover, are irrelevant. Further, establishments are required to name only the primary economic activity and to describe it in detail.

[3] In order to document this advantage in more detail, we used data generated by Schild (2016) who had matched establishments to legal entities (firms). The ten largest firms in Germany (each assigned to one NACE 3-digit code) had at the median 44.5 establishments. When aggregating the industrial activities in these establishments to NACE Rev. 2 3-digit codes, firms had (at the median) activities in 8 industrial codes.