# A Novel Technology-Industry Concordance Table Based on Linked Inventor-Establishment Data

Matthias Dorner (Max Planck Institute for Innovation and Competition, Munich; Institute for Employment Research, Nuremberg)

Dietmar Harhoff (Instutite for Employment Research, Nuremberg; Ludwig-Maximilians Universität (LMU), Munich; Centre for Economic Policy Research (CEPR), London)

Die FDZ-Methodenreporte befassen sich mit den methodischen Aspekten der Daten des FDZ und helfen somit Nutzerinnen und Nutzern bei der Analyse der Daten. Nutzerinnen und Nutzer können hierzu in dieser Reihe zitationsfähig publizieren und stellen sich der öffentlichen Diskussion.

FDZ-Methodenreporte (FDZ method reports) deal with methodical aspects of FDZ data and help users in the analysis of these data. In addition, users can publish their results in a citable manner and present them for public discussion.

## Contents

## Abstract

Mapping technologies into industries is frequently required in empirical innovation studies, but many concordances only provide coarse mappings. We develop novel concordance tables between industries and technologies making use of linked inventor-employee patent data for Germany. These data comprise 235,933 patents filed between 1999 and 2011 at the European Patent Office. Data on inventors are matched and disambiguated with social security records available at the Institute for Employment Research. Employment data recorded in this database include detailed industry codes describing the industrial activities of the inventors' establishments. The linked inventor-establishment microdata allow us to identify the precise industry of 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 concordance tables with existing work, and we discuss the validity of patent statistics by industries as indicators for innovation.

## Zusammenfassung

Die Verknüpfung von Industrie- und Technologiedaten wird häufig für empirischen Arbeiten zu Innovation benötigt. Die Verwendbarkeit existierender Konkordanztabellen ist jedoch durch deren Methodik bzw. Kompatibilitätsprobleme der Daten in der Praxis oftmals stark eingeschränkt. Dieser Methodenreport stellt einen neuartigen Ansatz zur Generierung von Konkordanztabellen vor, der auf verknüpften Erfinder-Betriebs-Daten basiert. Diese Daten enthalten Angaben zu 235,933 Patenten, welche von Erfindern in Deutschland zwischen 1999 und 2011 beim Europäischen Patentamt angemeldet wurden. Erfinder in den Patentdaten wurden mit Beschäftigten in Erwerbsbiografiedaten des Instituts für Arbeitsmarkt- und Berufsforschung (IAB) mittels Record Linkage verknüpft und disambiguiert. Aus der Kombination von industrieller Tätigkeit der Betriebe in denen Erfinder zum Zeitpunkt ihrer Patentanmeldungen arbeiten und den Angaben zu Technologien der angemeldeten Patente, lassen sich mittels Aggregation neuartige, auf Mikrodaten basierende Konkordanztabellen erstellen. Wir vergleichen unsere Technologie-Industrie Konkordanztabellen mit existierenden Ansätzen. Weiterhin wird die Verwendung von Patentindikatoren, die mit der vorgestellten Konkordanztabelle geschätzt wurden, als Maß für die Innovationsstärke von Industrien diskutiert.

**Keywords:** patents, international patent classification, industry classification, concordance, linked inventor-establishment data

**Download supplementary materials**

- Concordance tables (MR_07-17_concordance_tables, 7zip archive, 3 MB)
- Patstat appln_id to NACE industry (MR_07-17_appln_id_ipc4_to_NACE, 7zip archive, 170 MB)

**Acknowledgements:** We thank Reinhard Sauckel for his excellent research assistance. We gratefully acknowledge comments and suggestions received from Travis J. Lybbert, Nicolas J. Zolas, Giorgio Triulzi and Stephan Brunow, as well as from seminar participants at the Institute for Employment Research (IAB) in Nuremberg, at the Leibniz-University of Hannover and at the Max Planck Institute for Innovation and Competition in Munich. All remaining errors are our own. During this research, Matthias Dorner received funding from the Graduate Programme of the Institute for Employment Research (IAB-GradAB).

# 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 could 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. But prominent patent classifications (such as the International Patent Classification, henceforth: IPC) 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 level 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 exist already. 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 a concordance are based on firm-level information that were 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, their industrial activity is determined primarily from global value added (or turnover) based on the most important line of business. Since these large organizations hold the majority of patents, the precision of a concordance will typically be less than satisfactory if firm-patent linkages are used to construct the concordance. 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 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

[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).

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 concordance table. 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 inventor 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 find plausible similarities between our concordance and existing approaches documented in the literature. 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 in the supplementary appendix will hopefully allow 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. If these alternative data sources are unavailable, patent based indicators, which are less costly to obtain, provide a reasonable data substitute. The remainder of the

[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.

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 directly be converted 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 Putman (1997) used the YTC to predict patent counts by industry for the years 1983 to 1993. Results revealed that the predictions are fairly reliable for 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 aggregate manufacturing sectors) based on Finnish ISIC codes prepared by Statistics Finland. This concordance uses weights to link 4-digit IPC classes to these industry sectors. The quality of this particular concordance table is largely unknown, since it has not been employed frequently and since it has not been tested systematically.[4]

In 2002, Johnson provided an additional concordance between IPC and ISIC codes, referred to as the OECD Technology Concordance (OTC). Johnson (2002) used the IOO and IOU codes that were used as the basis for the YTC and translated them into the Canadian SIC system. To make the results compatible with international data, in a second step, the Canadian SIC system was translated into ISIC codes. Even though a novel link to ISIC codes was established by recoding of the industry classification, the OTC still suffers from the same problems as the original YTC.

Another prominent concordance, the so-called DG Concordance table, was constructed by Schmoch et al (2003). These authors developed an assignment of IPC codes to 44 industrial fields based on NACE industries. They identified a number of industrial sectors and their associated technological classes. The fit between the industry and technology classes was then thoroughly tested by investigating the patent activities of about 3,000 firms active in these industry sectors. This approach led to a concordance table which allowed translating industrial fields one-to-one into (dominant) IPC technology classes and vice versa. The resulting matrix was validated by comparing the patent data with export structures. Even though the team used a sophisticated approach combining expert knowledge with empirical data to determine and verify industry-technology links and further provides results that both seem to be internationally comparable and easy to use, there are two important disadvantages: first, the data used is based on firm-level industry codes and second, the authors link IPC classes only to manufacturing industries and services are entirely disregarded. The original DG Concordance table by Schmoch et al. (2003) constructed for NACE Rev. 1 data was updated in order to facilitate linking IPC codes with the current NACE Rev. 2 classification system (Van Looy et al. 2014). This revised concordance is also supplied with the regular updates of the PATSTAT data.[5]

Recently, a very different approach to link industry and technology data has been implemented by Lybbert and Zolas (2014). These authors employ a semantic matching technique which exploits automatic keyword identification and methods of text analysis. The text elements of descriptions of ISIC industry codes are compared to technical descriptions of patents, and measures of similarity are derived from this text data comparison. Since this probabilistic approach (Algorithmic Links with Probabilities, ALP) is implemented independently from empirical data on inventors or firms, which would require substantial efforts to process, it is

[4] Maurseth and Verspagen (2002) use the Merit concordance in their citation based analysis of knowledge spillovers in Europe. In a recent paper, Glitz and Meyersson (2017) use the concordance to estimate patent counts by industry for the GDR and West Germany in their analysis of the effects of industrial espionage on factor productivity differentials prior to German reunification.

[5] See https://www.epo.org/searching-for-patents/business/patstat.html (last downloaded on Aug. 8, 2017) for a description of the PATSTAT database).

relatively easy to update and adapt across countries with different industry classifications. With regard to the important conceptual distinction between IOU and IOO, by construction, this novel method should generate a concordance table that is particularly well-suited to the characterization of the economic context of use. Conversely, the approach proposed by Lybbert and Zolas (2014) should be less informative for scholars who are interested in the industrial origins of technologies.

The described concordances, while certainly helpful to researchers, have two important disadvantages. First, there has not been a systematic updating of the concordances using a proven and consistent methodology. This problem becomes evident as soon as industrial classifications are being revised. The methodology by Lybbert and Zolas (2014) circumvents this problem by abstracting from actual economic data. Second, the use of firm-level data in some of these attempts limits the precision of the concordance tables. In particular, in case of large multi-site and multi-product firms, the methodology used to determine industry codes constitutes another obvious weakness. The first problem can be tackled by employing an external data source for the construction of concordance tables that is updated reliably. Administrative employment data from social security records satisfy this requirement, and our approach described below lends itself to systematic replication and updating in the future. Utilizing administrative data also helps to lessen the second problem, since these data are usually organized by establishment as the reporting statistical unit in which industrial activity is more focused and more accurately measured than at level of a multi-product firm. Additionally, one has to conceptually distinguish between an IOO and an IOU approach. While the prominent DG Concordance and recent work by Lybbert and Zolas (2014) are certainly more appropriate to describe IOU relationships, the concordance proposed by us, however, is meant to deliver very precise locus-of-origin information.

## 3 Data and methodology

The technology-industry concordance tables proposed in this paper are generated from linked inventor-establishment data that contain detailed technology class information from patents and establishment level industry codes originating from administrative employment data. Data of this kind allow us to link the technology classes from patents with the establishments where inventors generated these inventions. Thus, our approach towards constructing a concordance between industries and technologies is entirely based on administrative data with a high level of precision, reliability and regular updates.

The patent data sample is based on filings at the European Patent Office (EPO) between 1999 and 2011 that are recorded in the PATSAT database. We excluded all applications solely filed by foreign entities, since no data on establishments located abroad are recorded in German administrative data. For the same reason we also excluded inventors with residential address outside of Germany. The resulting raw data sample contains 293,145 patents (699,894 patent-inventor records). The corresponding inventor data, i.e. names and addresses reported on the patent document, were pre-processed and subsequently matched with administrative

employment data using methods of record linkage.[6] These administrative employment records were generated in the social security system and cover the universe of employees in Germany with the exception of civil servants (incl. regular professors) and self-employed workers. According to the PatVal Survey, these groups account for about 10 percent of all patent filings at the EPO (Gambardella et al. 2005). The employment data are stored, processed and prepared for scientific use as linked employer-employee data by the Institute of Employment Research (IAB). For the purpose of record linking, we used name and address data on employees that can be matched with the inventor information recorded on patents. We relied on both deterministic as well as fuzzy matching methods to determine links between inventors and employees.[7]

The matched sample obtained from the record linkage is comprised of 148,743 unique inventors respectively employees. These individuals were involved in the filing of 235,933 patents between 1999 and 2011, corresponding to 524,386 matched patent-inventor records. At the patent level, the matching rate equals 80.5 percent, i.e. for this fraction of the raw patent data at least one inventor from the team listed on the patent was matched unambiguously to an employee.[8] For 60 percent of the patents (140,577) comprised in our matched data set, we were able to link the full team of inventors in the patent data, while the remainder contains only patents with one or more inventors who could not be matched. Computations and the concordance tables presented below are based on the matched sample including all patents, but we performed robustness tests based on a data subsample that was restricted to completely matched inventor teams.

To reference patents unambiguously with the employment episodes of the matched individuals we select the employment episode in the linked employer-employee data of the IAB at the time of the patent filing date recorded in the PATSTAT database. If an employee holds multiple parallel jobs with different employers at this point in time, we restrict the employment data to the main job, i.e. the employment record with the highest wage subject to social security as reported in the IAB data.[9] Patent filings that intersect with episodes of unemployment or episodes without register information in the labor market biography data are discarded in our analysis since they do not provide industry codes from any employer. In total, inventors were employed in 23,073 different establishments when filing patents between 1999 and 2011. Inventor teams listed on a patent may be employed with the same establishment or in different

[6] Preprocessing included cleaning of the strings, parsing and extensive checks (e.g., deletion of corporate inventors or deletion of erroneous addresses). For the purpose of record linking we had access to confidential data on the names and residential addresses of more than 30 million employees in the German social security system in each year. The record linkage was conducted at the Research Data Center (FDZ) at Institute for Employment Research (IAB) in Nuremberg.

[7] A detailed description of the methodology used in the pilot study of this work is provided in Dorner et al. (2014).

[8] Given that civil servants and self-employed workers are not recorded in the IAB data and we could not identify their patent filings in the initial patent data, the reported matching rate represents a lower bound of the actual quality of the linkage.

[9] The IAB data record 2,140 patent-inventor observations with parallel employment episodes in different establishments. This number equals a fraction of 0.4% of all matched records in our database and relates to 728 inventors. Using the NACE Rev. 2 divisions (two-digit numerical codes) we find that the equivalent of 11.5% of the patents in the sample are filed by inventors working for establishments that operate in different industries (mean 1.12; std. dev. 0.37; max. 6).

establishments.[10] However, differently from applicant data recorded on patents, the number of establishments cannot exceed the number of inventors.

To generate technology-industry concordance tables from these linked data we require classifications for both industries and technologies. Technology classifications are derived directly from the patent data. Technology codes in the IPC are assigned by the patent examiners to describe the technology of patented inventions and facilitate the search for prior art and related technology. To obtain a unique technology classification from the set of available IPC classes (4-digit level) listed on a patent, in the first step, we recoded the very detailed IPC data into a slightly modified version of the technology area classification proposed by Schmoch (2008) (see Table A1). We use the reclassified technology information to compute a dominant technology area for each patent as defined by the modal value.[11] Correspondence tables are generated for different aggregations of technology areas consisting of 30, 34 and 35 categories. We further computed a version of a concordance between IPC classes (4-dig.) level and industries in which we use fractional weights for industries as well as for technologies (see section 5).

The precise link between inventors and their employment episodes gives us access to detailed industry codes from the IAB data. Industry codes are recorded in the NACE classification system that describe the main industrial activity of the establishment at the time of the patent filing and are available in the precision of the 5-digits level.[12] Thus, industry information that are available to us are much more closely related to the inventor and the economic activities at the origin of the patent itself, than patents linked to the economic activity of a whole company or firm that are usually determined by statistical agencies on the basis of sales data. An important issue in many of the existing concordances and industrial analyses is that industry classifications outdate after some years because of updates. The NACE classification system was subject to two major updates between 1999 and 2011. The first (minor) update from the NACE Rev. 1 to the NACE Rev. 1.1 occurred in 2003, while the latter classification was replaced by the redesigned NACE Rev. 2 system in 2008. We followed the methodology proposed by Eberle et al. (2011) and generated time consistent NACE classifications (Rev.1, Rev. 1.1, Rev. 2) from the IAB population data for all establishments recorded in our matched database.

[Figure 1 about here]

[10] The administrative employment data does only provide a unique establishment identifier but no additional information that indicates whether different establishments belong to the same firm.
[11] If the modal value of the technology classes was ambiguously defined for a patent, we took a random choice from the possible candidates. We inspected the full range of IPCs as well, and while there is considerable variation at the level of full-length IPC codes, much of that vanishes after aggregation to technological areas. Concordance weights obtained from a fractional methodology are highly correlated with dominant technology class weights for each patent.
[12] The exact regulations that apply to establishments reporting their industry activity to the social security authorities are defined by the Federal Employment Agency (BA). Guidelines of the BA require establishments to report their industry code to the social security authorities (at least) at an annual basis and based on the main objective of the firm, i.e. the core economic activity of the majority of employees in the firm. Sales or other business accounting indicators that are used for the same purpose by statistical agencies are irrelevant.

Figure 1 depicts two exemplary inventor biographies (inventor I1 and inventor I2) and related patents (P1, P2 and P3) with their representation in our linked data.

The actual industry-technology concordance tables generated from our data are based on (fractional) counts of co-occurrences of industry codes *i* and technology area *t*. Records considered for the concordance tables include only patent-inventor records that overlap employment spells such as depicted in Figure 1 for the patents P2 and P3. Patent P1, however, refers to a patent that was filed during an unemployment period of inventor I1 and thus does not include industry codes from any establishment (see Figure 1, upper left graph).

There are two ways to generate the industry-technology concordance tables: the first approach would be to count unweighted matched patent-inventor observations as indicated in column *inv* of the table documented in Figure 1. The resulting sum over industries and technologies then includes multi counts of each patent times the number of matched patent-inventor records. The preferred approach, however, relies on weighted patent-inventor records, i.e. inventor fractional counts.

$$frac_{inv} = \frac{1}{N_{inv}}; with\ N_{inv} = unique\ inventor - employee\ matches\ per\ patent$$

In this representation patent-inventor records are assigned equal fractions that relate to the size of the matched inventor team and each patent is counted by the value of unity. The fractional counts are reported in column *frac_inv* of the data table in Figure 1.

Collapsing the inventor fractional counts by industry and technology yields a two-way contingency table where the number of rows corresponds to the technology areas and the columns correspond to the number of industries, respectively. The data cells in the table record frequencies of industry-technology co-occurrences ($y_{it}$). In both dimensions of the table we can compute the marginal row ($Y_{t.}$) or column frequencies ($Y_{.i}$) by simply calculating the totals for rows and columns.

$$Y_{t.} = \sum_{t=1}^{T} y_{it}\ ;\ with\ T = Number\ of\ technologies\ (columns)$$

$$Y_{.i} = \sum_{i=1}^{I} y_{it}\ ;\ with\ I = Number\ of\ industries\ (rows)$$

Substituting the fractional counts ($y_{it}$) with the relative frequencies per row ($\frac{y_{it}}{Y_{t.}}$) or column ($\frac{y_{it}}{Y_{.i}}$) yields our industry-technology concordance table, where in the horizontal or in the vertical dimension of the table, the totals of the relative frequencies sum up to the value of unity:

$$\text{Horizontal table: } y_{t.} = \sum_{t=1}^{T} \frac{y_{it}}{Y_{t.}} = 1$$

$$\text{Vertical table: } y_{.i} = \sum_{i=1}^{I} \frac{y_{it}}{Y_{.i}} = 1$$

In the horizontal table, the relative row frequencies ($\frac{y_{it}}{Y_{t.}}$) indicate the fraction of inventors working in industry *i* who contribute to patent filings in technology area *t*. In the vertical table, the relative column frequencies ($\frac{y_{it}}{Y_{.i.}}$) document the fraction of inventor's patent filings in technology *t* which originate in industry *i*.

# 4 Technology-industry concordance tables

## 4.1 Descriptive statistics

In this section we present and discuss an exemplary technology-industry concordance table. As a technology classification, we build on 34 technology areas (TF34) aggregated from the IPC codes following the proposal of Schmoch (2008).[13] Industry information is presented in the NACE Rev. 2 classification system and at the 2-digits level describing 86 unique industry divisions. Before we present our concordance table, we describe relative frequencies of both, technology areas and industry divisions in our matched sample. Table 1 reports the technology areas.

[Table 1 about here]

Overall, the empirical distribution based on the matched sample represents the technology portfolio of the Germany economy with its strength in moderate technology intensive sectors such as machinery, automobiles and electrical industry. Technology intensive high-tech sectors such as ICT or pharmaceuticals are underrepresented when comparing the technology shares to the U.S. (Schmoch and Frietsch 2010). High correlations of the shares with the respective full population data on patent filings in Germany ($\rho > 0.9$) indicate that the matched subsample and its composition is representative for the national technology portfolio (see Table A3).

In quantitative terms, '31 Transport' is the most frequent technology area in our linked data. The share of patents filed in this particular technology amounts to 8.7%according to totals of inventor fractions. '1 Electrical machinery, apparatus, energy' and '30 Mechanical elements' follow in the ranking of the most important technology areas with patent shares of 6.9% and 5.5%, respectively. At the bottom of the ranking we find fields such as '5 Basic communication processes', '11 Analysis of biological materials', '18 Food chemistry' and '7 IT methods for management' which account each for less than 1% of the patents in our data set.

A unique feature of our linked data is that we are also able to directly identify the precise industrial origin of patents recorded in our data set via the inventor (see Figure 1). Using a

[13] TF34 is a slightly modified version of the originally proposed TF35 classification by Schmoch (2008). The only difference is in the aggregation of technology fields '21 Surface Technology' and '22 Nanotechnology' into one joint technology area.

(fractional) count based on our inventor-establishment data and patent population shares, we document the distribution of industries in our sample in Table 2.

[Table 2 about here]

It is evident that industries contribute differently to the national patent output of Germany, however, as expected, patenting is also highly concentrated on a small group of industries with both technological opportunities and a market environment that make them more likely to use the patent system. Ranking the industries according to their quantitative contribution to overall patenting, the cumulative share of the ten top ranked branches exceeds already 75%of the patent population. For the median industry, the respective value is already 98%. When focusing on the industries, the highest number of patents in Germany originate in the industry division '28 Manufacture of machinery and equipment n.e.c.'. The patent share of this top ranked industry amounts to 14.6%, which is disproportionately high compared to the industry's share in total employment of only 2.8%in 2010. '26 Manufacture of computer, electronic and optical products' is ranked second with a patent share of 12.2% and '20 Manufacture of chemicals and chemical products' is the third largest industry of origin of patents (10.7%). Our approach also documents the contribution of service sector establishments to patenting. '72 Scientific research and development' is ranked 5$^{th}$ (8.7%) and '71 Architectural and engineering activities; technical testing and analysis' is ranked 9$^{th}$ (2.9%). The fact that two large service branches are represented among the top patenting industries highlights that service divisions are also important contributors to the technology portfolio and that omitting them in concordances introduces bias in estimates of industry-technology relationships. Actually, the share of inventors being employed in service sector establishments is actually larger than previous research has suggested (see Blind et al. 2003). A correlation analysis of industry level shares in patenting and corresponding industry level employment shares computed from administrative employment data of the IAB[14] shows that patenting is – as expected – positively correlated with industry shares of the science and engineering workforce ($\rho$ = 0.52). The correlation of industry level patenting and employment shares of the industries in the total workforce, however, is relatively low ($\rho$ = 0.17).

## 4.2 Concordance tables

We present two versions of the technology-industry concordance table between 86 NACE Rev. 2 divisions and 34 technology areas.

*a) Horizontal correspondence table*

Figure 2 depicts the horizontal table as a "heatmap" with industries shown on the X-axis and technology areas reported on the Y-axis.

[Figure 2 about here]

[14] All industry level indicators are derived from the IAB Establishment-History Panel (BHP). Employment shares refer to a pooled industry-panel data set covering 86 NACE rev. 2 industries over the period 1999-2011. The BHP database is discussed in the empirical application in greater detail.

The data cells in Figure 2 are colored according to their relative weight to total patenting in a single technology area (Z-axis). With increasing weight of an industry to patenting it is filled with a darker color, i.e. the darkest colored cell in each row represents the dominant industry in a technology area. Overall, it is evident from the heatmap that most technologies tend to have a strong dominant industry that contributes the lion's share to patenting in a technology field. The strongest relationships are usually found among manufacturing industry divisions but also in knowledge intensive branches of the service sector.

The dashed rectangles depicted in Figure 2 are used to exemplify how to read the tables. The first technology area highlighted in the table is '6 Computer technology'. The most important contributor to patents in this particular technology area is the industry '26 Manufacture of computer, electronic and optical products', which accounts for roughly one third of the patents (33.8%). The remainder of the distribution is exemplary for general purpose technologies (GPT) that require a set of diverse industry inputs. '62 Computer programming, consultancy and related activities' is ranked second with a share of 16.9% and highlights the importance of software development for patents in this technology area. Further important industries with notable contributions are '72 Scientific research and development' (11.0%), and '27 Manufacture of electrical equipment' (6.1%). While the latter is related to hardware components developed in the electrical industry, the former indicates that knowledge from academia is essential in this technology.

The technology field '17 Polymers' differs substantially from the GPT example above as its industry inputs are highly concentrated. Patents in '17 Polymers' originate almost exclusively in the industry division '20 Manufacture of chemicals and chemical products' (78.9%) with some minor shares in '72 Scientific research and development' (4.9%) and '22 Manufacture of rubber and plastic products' (3.5%).

The two examples highlight that technologies differ strongly in their industry specific inputs. These differences across technology areas can be summarized by a Herfindahl-Hirschmann index (HHI) computed over the industry fractions.[15] The mean HHI is 0.213 (std. dev. 0.103) indicating moderate concentration on average. '17 Polymers' is the technology area with the maximum HHI value (0.621), while'6 Computer technology' ranges among the technologies with the lowest concentration (0.166). The minimum value of the HHI (0.098) is found for '21 Surface technology, coating'.

*b) Vertical correspondence table*

The second version of the concordance table, the vertical structure, is depicted in Figure 3. The dimensions of the figure conform to the horizontal structure and also the interpretation of the color scale is analogous to Figure 2, however cell values now indicate the weight of technology output in a specific industry, i.e. in which technology areas inventors in a single industry file their patents.

[15] In the innovation literature, this indicator is also interpreted as a measure of generality of patents (Bresnahan and Trajtenberg 1995).

[Figure 3 about here]

Consider the example of the ’20 Manufacture of chemicals and chemical products’ which is highlighted in Figure 3. We expect that inventors employed in this industry predominantly patent in related technologies. This is confirmed by the profile of the patented technologies. The most important technology originating from this industry division is the technology area ‘14 Organic fine chemistry’ with a share of 23.9%. Other technologies with substantial shares in this industry are the technology areas ’17 Polymers’ (20.6%), ‘19 Basic materials chemistry’ (16.2%) and ‘22 Chemical engineering’ (6.8%). Inventors employed in the R&D units and in academia (’72 Scientific research and development’), however, patent a more diverse set of technologies. Here the focus is on science based technologies as opposed to engineering driven technologies, which tend to be more prominent areas of commercial patenting. The highest share of patents originating in this particular industry is filed in the field ’15 Biotechnology’ (11.8%). ’14 Organic Chemistry’ (8.7%) and ’16 Pharmaceuticals’ (8.3%) follow in the ranking.

Using the HHI as an indicator for the concentration of technology output of industries we find an average concentration of 0.162 (std. dev. 0.123). Thus, the average concentration in the vertical dimension is less pronounced in the vertical than in the horizontal structure. The two examples of the chemical industry and the academic sector are both examples for a low sectoral concentration of technology production (20: HHI = 0.138; 72: HHI = 0.056). The minimum of the HHI is 0.042 (’84 Public administration and defense; compulsory social security’). The industry with the highest concentration (HHI = 0.542) is found in ‘12 Manufacture of tobacco products’, where 71 percent of the patents are filed in technology area ’33 Other consumer goods’.

# 5 Tests of the technology-industry concordances

We provide several empirical tests in order to document potential advantages of our approach to generate concordance tables and to support potential users in making an informed choice between existing concordance tables for their empirical application. First, we test the concordance tables proposed above on the presence of potential matching bias resulting from incompletely matched inventor teams (5.1) and on temporal variation of the underlying technology-industry relationships (5.2). We then analyze the differences between our concordance and two existing works, the ALP and the DG concordances (5.3). Toward this end, we exploit the flexibility of our approach relying on linked inventor-establishment data to generate structurally compatible concordance tables based on IPC classes (4-digit) and sectoral aggregations of NACE industries. Third, we use our concordance table presented above to compute patent counts by industry for Germany and evaluate patent intensities of these industries as an indicator for innovation output against alternative measures derived from comprehensive survey and administrative data (5.4).

## 5.1 Matching bias

The inclusion of incomplete inventor teams, i.e., patents in which not the full number of inventors were matched, could potentially bias the weights reported in our concordance table

because unmatched inventors could work for establishments which operate in different industries. We test for the presence of this bias by comparing the concordance table based on the full sample (235,933) with a second concordance derived from a reduced patent subsample that contains only completely matched inventor teams (140,577). When we compare the modal weights and classes of the concordance tables we find a high degree of conformity. In the horizontal structure, the modal technology areas are equal in 33 out of 34 (97%) technologies and in the vertical table in 60 out of 86 (70%) industry divisions. Further, the actual modal values are also highly correlated (horizontal: $\rho$ = 0.98, vertical: $\rho$ = 0.86), indicating a high degree of similarity in the core properties of the concordance tables.

Extending the test to the full concordance matrices we find that cell weights – as long as they are represented in both tables – are highly correlated between the two versions (horizontal: $\rho$ = 0.99, vertical: $\rho$ = 0.89). Overall, there are only minor differences between the concordance based on the full sample and an alternative version that is based on patents that contain only completely matched inventor teams and thus is corrected for bias related to unmatched inventors in record linkage.

## 5.2 Variation of the concordances over time

In the second test, we investigate the variation of the concordance table (full patent sample only) over time by comparing the tables generated for the years 2000, 2005 and 2010 among each other and with the pooled version presented above.

First we test the modal classes of the different concordances for the same industry division and technology area, respectively. In the horizontal table we find that in 21 out of 34 (61.8%) technology areas the modal industry division conform in all years with the pooled concordance table. The corresponding value in the vertical structure is only 30.4 percent, indicating that industries adjust their most important output technology over time.

We now extend the test of the temporal robustness to the full matrix and compute correlations of the cell values over time. Note that correlations are only computed if an industry-technology combination is present in all years. Thus, this analysis considers mainly the core of the correspondence while it may miss (temporary) niches in the technology-industry relationship. The full correlation matrix that is based on 2,346 out of 2,924 possible co-occurrences is depicted in Table 3.

[Table 3 about here]

The correlations reported in column 1 between the pooled table and the data cells in the annual correspondences indicate high persistence of the industry-technology relationship over time in the horizontal version. In the vertical table, correlations between the annual tables and the pooled version drop but still exceed the margin of 0.6. We can slightly increase the correlations documented in Table 3 by restricting the sample to the top five ranked industries.

Generally, it is evident from all analyses in which we compare the annual correspondences with the pooled table, that the year 2010 is different. These differences in the correspondence for 2010 are most likely related to the global recession in the years 2008/2009. The recession

may have affected patenting decisions in industries that were disproportionally hit by the recession. This could have led to a relative shift in patented technologies of some industries. For a more detailed analysis of the persistence of a technological shift, however we would require a longer time series, which is currently not available.

## 5.3 Comparative analyses

### 5.3.1 ALP concordance (Lybbert and Zolas 2014)

We discuss differences between our concordance and the ALP concordance proposed by Lybbert and Zolas (2014). To this end, we compute from our data a concordance table that conforms to the structure of the ALP concordance, i.e. it maps fractionally weighted IPC4 classes, instead of the dominant 34 technology areas, into NACE Rev. 2 industry divisions (and vice versa). An issue of earlier work besides the updates of classification systems was also the incompatibility of industry classifications across countries. Here, we benefit from the harmonization of NACE Rev. 2 and ISIC Rev. 4 classifications, which correspond to each other at the level of NACE divisions and 2-dig. ISIC industries.

The first and most obvious difference between our and the ALP concordance is the industrial coverage. The ALP concordance assigns only manufacturing industries (NACE Rev. 2, divisions 10-42) to IPC4 classes (and vice versa). This restriction is related to the industry of use nexus that is described by technologies and industrial descriptions used in the matching. Holding the set of industries constant across the tables, it becomes evident that our concordance covers a larger number of industry-technology relationships than the ALP concordance, which is restricted by the keywords in the industrial profiles. This advantage with respect to a more detailed representation of the subject is also reflected in the average number of IPC4-industry links, which is significantly higher in our concordance compared to the ALP version (IPC4 to NACE Rev. 2.: 25.24 vs. 3.78; NACE Rev. 2 to IPC4: 143.60 vs. 22.33).

A more focused comparison concentrates on the most important IPC4-NACE-Rev. 2 relationships across both concordances in terms of the relative weights. Correlations at the level of data cells with non-missing weights and ranks are presented in Table 4.

[Table 4 about here]

First, the analysis of the relative rank orders in the overlapping sample of the two concordances shows that in at least 50 percent of the cases, the top ranked IPC4 to NACE industry (and vice versa) link are the same. These overlaps increase substantially for the horizontal table (column 2) if the equality of the top ranked NACE industry or IPC4 class is tested for being listed among the top three ranks of the corresponding ALP table.

Second, the analysis of correlations of the actual weights yields mixed results. While the rank orders are better preserved in the case of the horizontal table (column 2), the actual weights appear as moderately correlated ($\rho = 0.373$). This finding is maintained also for subsamples in which the overlap is restricted to the top ranked links only. The opposite is actually found for the vertical table (column 1). While the ranks are less well preserved across the tables, for equally ranked IPC-NACE relationships, however, the weights appear as being more similar

($\rho$ = 0.630). Generally, this is a quite encouraging result especially for users of the ALP concordance who find their top ranked industry-technology links also represented in empirical data, as well as with respect to IOO and IOU comparisons.

The average differences in the weights amounts to about 12 percentage points in the horizontal table, a quite substantial figure, and only 3 percentage points in the vertical table, respectively. These differences reflect the significant variation found in the correlation of the weights.

[Figure 4 about here]

Figure 4 presents kernel density plots for the cell based differences in the weights between the two concordances. It is evident that the differences are normally distributed, with the majority of the data points differing only marginally across the concordances. Nevertheless, the especially the long tails in the negative part of the distributions of the differences indicate that many ALP weights are significantly higher than their empirical counterparts in our concordance. Since raw empirical weights are used in our concordance as opposed to more sophisticated weighting approaches in the ALP concordance, reweighting might further help to reduce these differences.

### 5.3.2 DG Concordance table (Schmoch et al. 2003)

The DG Concordance table is one of the most popular tools for the purpose of connecting technology with economic data. It links technologies to industries based on a one-to-one mapping of 4-digit IPC groups into 44 different manufacturing fields, defined by NACE industry divisions. Given the structural differences between the approaches towards generating concordance tables (one-to-one vs. weighted links), we provide two simple tests of differences. To this end, we computed from our data a customized concordance table using the 44 manufacturing fields (NACE Rev. 1.1) and IPC4 classes.[16] Since the DG Concordance is mainly used to map patent counts by IPC4 classes into industrial data (see e.g., Danguy et al. 2014), we compare it only with the horizontal structure of our concordance. We first analyze the overlap between technology-industry associations recorded in our concordance and those documented by Schmoch et al. (2003). In a second step, we compute correlations between the (average) weights in our concordance and the (pre-determined) weight of unity in the DG concordance.

Comparing our customized concordance with the DG version, again, highlights the advantage of our concordance table with respect to detail.[17] Our table includes 11,162 weighted links between NACE fields and IPC4 classes, while the DG Concordance is limited to only 615 (one-to-one) records. The merged sample represented in both tables includes 558 data cells, covering 93% of initial DG Concordance but only 5% of our initial concordance table. These 558 records, however, appear to represent the core records of our concordance. We draw this conclusion from the significantly higher average weights of these data cells, as compared to the unmatched fraction (matched: 0.276; unmatched: 0.030, $p<0.01$). Another way to contrast

[16] Instead of assigning a dominant technology to each patent (see section 3), we used a weighted approach in which IPC4 classes were assigned frequency based weights at the patent level.
[17] Note that this comparison is limited to the 44 NACE fields and neglects the other 1,056 industry-technology linkages in NACE divisions that are omitted by Schmoch et al. (2003).

the two concordances is by simply counting the average number of links within the respective industry/technology classes being used. While Schmoch et al. (2003) assign each IPC4 class only one single industrial field, we document for the 633 IPC4 classes on average 18 NACE fields. The respective average number of IPC4 classes assigned to the NACE fields amounts to 14 in the DG Concordance as opposed to 254 in our concordance.

The overall correlation of the weights in the two concordances amounts to 0.495.[18] This is actually a remarkably high statistic given the significant differences in the methodologies. Nevertheless, there is also strong variation across NACE fields underlying this global statistic. While there are some NACE fields in which the weights of the two concordances are highly correlated, i.e. the NACE field weights in our concordance are often close to the value of one, as defined in the DG Concordance, in others the concordances differ quite substantially. The fields in which similarity is the highest are '31 Accumulators, Battery' ($\rho$ = 0.911), '23 Agricultural and forestry machinery' ($\rho$ = 0.794) and '40 Optical instruments' ($\rho$ = 0.776). These industries have in common, that they link only few IPC4 classes in the DG Concordance to the respective industrial field. In our concordance table, however, the number of links is substantially higher as expected from the use of more detailed establishment data. Moreover, in these cases also the benefits of a weighted approach as opposed to singleton links become obvious. NACE fields in which the correlations of the IPC4-to-NACE field relations across tables are weak include '33 Other electrical equipment' ($\rho$ = 0.139), '39 Industrial process control equipment' ($\rho$ = 0.202) and '3 Textiles' ($\rho$ = 0.243). For the field '12 Paint, varnishes' we even find a slightly negative correlation of -0.015. For these fields in particular it is very likely that the mapping of patents by technologies into industries yields very different results. The same disclaimer applies to the vertical table. While our concordance table still shares the advantages in terms of a more detailed depiction of industry-technology relationships, correlations of the weights with their equivalents (unity) in the DG concordances drop significantly ($\rho$ = 0.260).[19]

### 5.3.3 Discussion

Despite the differences in the underlying data and methodologies used for generating the concordances, we find that the key technology-industry relationships are similar across the three concordances subject to this comparative analysis. Nevertheless, encouraging overall correlations of ranks or actual weights in a single statistic appear to hide significant heterogeneity across industries and technologies, respectively. To this end, users of each of the concordances might carefully reconsider and reflect on their results in the light of these differences.

Most likely, the flexibility of our approach will help users to find an adequate concordance that is suited for their specific data needs and industry subsamples. Certainly, however, our high-quality data and methodology are superior with respect to capturing details and niches of actual technology-industry relationships. To this end, the ALP concordance has to rely on more or

[18] NACE-IPC4 links that are only represented in one of the two concordances, are recoded from missing to the value of zero, respectively. This yields a balanced sample of 11,219 records with weights [0;1].
[19] Detailed results of the comparative analysis will be provided by the authors upon request.

less arbitrary cutoffs that determine the scope of their concordance and the approach used for the construction of the DG Concordance structurally neglects heterogeneity within firms by restricting the relationships with technologies on the primary of industry of operation of the approx. 3,000 firms in the sample.

Nonetheless, this increase in detail might come at the cost of including casual relationships or artefacts of technology-industry co-occurrences. These records are represented in particular in the tails of the distributions underlying our concordance tables. However, using a weighted approach, as opposed to singleton linkages potentially also cures these issues as it provides users with the flexibility to define their own cutoff values based on weights, ranks or combinations of both. Hence, our approach exploits the advantages of comprehensive high-quality empirical data and a weighted approach. Since relative weights determine the results of any application of the concordances, users should also carefully consider the choice of horizontally versus vertically structured tables. While these tables are per definition the same in the DG Concordance, ALP and our approach produce two versions. Given that our concordance by its very concept and its empirical basis reflects the revealed technology choices of inventors in a given industry, researchers might therefore prefer our concordance over the ALP alternative especially in the case of knowledge production applications. Since the DG Concordance is based on a similar logic and empirical data, the correlations with our concordance are substantial for this particular version of the table. Nevertheless, in the light of issues pertaining to the determination of firm level industry codes, users should consider the context of their empirical analysis with respect to IOO vs. IOU. Differences in the vertical structure of the ALP and our table are only marginal so that users of the concordances might chose the concordances based on their specific data needs. To this end, again, our concordance should be especially appealing to those users who prefer a clear industry of origin view in their analysis and who either want to maximize coverage of their data and work with (disaggregated) data for the service sector.

## 5.4 Patents as indicators for innovation at the industry level

In this application, we assess as to whether the estimated numbers of patents by industries generated using our concordance are correlated with frequently used indicators of the innovation performance of industries. To this end, we use raw counts of national as well as a subset of transnational patents by technology area and priority year for Germany as obtained from PATSTAT.[20] We map these counts into patents by NACE Rev. 2 industry divisions using our concordance table (horizontal table, see Figure 2). The resulting industry-year panel data on patent counts are normalized with the size of industries, measured in thousands of employees according to social security data of the IAB. These data are complemented with other industry level innovation indicators computed from survey data. The results of the correlation analysis are presented in Table 5 (see Table A3 in the appendix for summary statistics). In Table 5, we show the bivariate correlation coefficients of the patent counts with indicators derived from both, administrative and survey data of the IAB data as well as industry level data available from the ZEW Innovation survey data.

[20] The definition of transnational patents was adopted from Schmoch and Frietsch (2010).

[Table 5 about here]

The correlation analysis at the level of 86 NACE Rev. 2 divisions shows mixed results. Generally, for the IAB data, we find that the magnitude of correlations between patents and other innovation indicators are significantly stronger in the manufacturing subsample, as compared to the full sample that includes also service and public sector divisions. Manufacturing sectors have higher R&D intensities, and as a result of scale economies, R&D inputs are used more effectively. Differences in appropriability further contribute to these stable lines of sectoral segmentation. Moreover, differences between national and transnational patents have only marginal effects. Among the correlations reported for the IAB data in Table 7, the most significant relationship appears between the share of science and engineering (S&E) workers and the patent intensity of industries. The share of highly skilled workers, as the superordinate group of S&E workers, is highly correlated with patenting intensity. Especially for manufacturing industries, also the share of firms that successfully introduced a new to the market (product) innovation appears as another indicator that is strongly correlated with the patent intensity of industries.

Columns five and six report the correlations for the two types of patent definitions with industrial indicators obtained from the sectoral reports of ZEW Innovation Survey[21]. These industries represent NACE Rev. 2 divisions as well as further aggregated sectors of manufacturing divisions and private sector service branches. While we find virtually no differences to the IAB data with regard to the correlations between the two groups of employees and patent intensity, the results on the other innovation indicators from the ZEW data are highly informative, as they show substantial correlation with patent intensity. With the exception of the share of innovators, which by definition includes firms who are active in process or organizational innovation of which both are hardly patentable, all relationships exceed the magnitude of correlation found for the share of science and engineering staff. The share of firms with continuous R&D activities is the indicator with the strongest correlation, nevertheless, also the measures based on product innovations (share of firms, share of turnover) as well as innovation expenditures (innovation intensity) are highly correlated with our estimates of patent output. Since the focus of the ZEW survey data is on innovation in manufacturing, the findings are in line with the previous results based on the IAB data, which, however, covered a larger scope of industries.

Overall, the correlation analysis highlights that patent counts and intensities generated using our novel concordance table provide both a convenient and valid solution to compute novel statistical indicators that approximate the patenting performance or innovativeness of industries. We also confirm that the correlation between patents and commonly used innovation indicators is particularly high for industries in the manufacturing sector. Empirical evidence of this kind is informative for applied researchers because patent count estimates by industry can substitute for the lack of data about innovation at the industry level, since survey data are often unavailable or hard to compare across countries. Moreover, since patent register data cover long time spans and are comparable across countries, especially cross-country

[21] These industry level data are publicly available from the ZEW website (http://www.zew.de/en/publikationen/zew-gutachten-und-forschungsberichte/forschungsberichte/innovationen/zew-branchenreport-innovation).

panel analyses can benefit from the use of concordances to generate industry level indicators of innovation based on patents.

## 6 Conclusions

This paper addresses the missing link between industry and technology data, which is of great importance for economic analyses of growth, innovation and technological change. The novel technology-industry concordance tables we describe are based on linked inventor-establishment data, combining patent register information from the PATSTAT database with administrative employment data originating from the German social security system. Inventive activity related to a particular technology can thus be observed directly in the organizational environment the inventor is working in – the establishment. Establishment level industry codes in administrative employment data are determined from the economic activity of the actual local site and are considerably more precise and fine-grained than those of (multi-site) firms. In the case of patenting, we argue that the use of establishment level data should give a concordance table substantial advantages in terms of precision over concordances based on industry classifications of firms. Per definition, these industrial activities recorded in firm data are dominated by the firm's most important line of business, and technologies or R&D activities in niches may not be recognized and identified correctly. Given that the lion share of R&D and inventive activities are concentrated in large multi-site firms of this kind, the availability of information at the establishment level should be helpful in avoiding aggregation biases.

Another significant advantage of our approach over existing concordances that use empirical data is its potential for systematic updating. Both data sources combined in our linked data set originate from administrative procedures and are updated on a regular basis with the important premise of inter-temporal reliability and harmonization. Thus, our methodology documents also a promising avenue to investigate the dynamics of industry-technology relationships in the future in greater detail. Over the period covered by our data, we find evidence for high persistence in the relationships captured by our concordance. However, recessions or technology shocks may impact on industry-technology patterns.

Our focus is to generate a concordance between the industrial locus of origin and the patented technology. We argue that our concordance should lead to more precise estimates than crosswalks derived from other data, especially in the context of knowledge production. Descriptive analyses confirm this presumption. We find some differences to existing concordance tables that have the potential to affect results of empirical analyses. Nevertheless, in the core features of the concordances, we find similarities that are preserved across the different methodologies. Our unique data enable us to extend technology-industry concordances towards the service sector that was systematically omitted by earlier work. To this end, we show that a significant share of inventive output is actually being generated in a service context and that technologies show also a rich variety of industry input patterns. This holds even more for the technology fields chosen by inventors employed in the same industry.

We admit that researchers might find our approach less informative about the locus of use of the invention. Another potential caveat of our approach is that we generate our concordance

tables based on German data. Earlier empirical work documented that the application of concordances across different countries might introduce bias. While we cannot fully solve potential critique in this regard as a result of data availability, we feel optimistic about the results obtained from comparative analyses against existing concordances, which were constructed from cross-country data. Notwithstanding this evidence, we encourage researchers to apply our concordance table also in their cross-country analyses of innovation and document potential deviations as compared to alternative approaches in their studies. While we see the main utility of our tool in the context of cross-country studies of industries and technologies, the very accurate and representative depiction of technology- industry relationships in Germany, one of the most active patenting countries in the world, will be also helpful for many case studies and empirical applications that have a focus on the national innovation system of Germany.

The concordance tables presented in this paper and tables for other combinations of technologies and industries are available from the supplementary appendix of the paper. [22] We further provide a table including both technology information (modal class ipc4) and NACE industries at the 3 digits level for all patents recorded in PATSTAT (version April 2017). Using the appln_id as merging key, NACE industry codes from these data may be directly merged with patent data sets based on PATSTAT.

## References


Aghion, P., Akcigit, U. and Howitt, P. (2014). What Do We Learn From Schumpeterian Growth Theory?. In: Aghion, P. and Durlauf, S. N. (Eds.), Handbook of Economic Growth, Edition 1, Volume 2, 515-563. Elsevier, Amsterdam.

Aghion, P., Akcigit, U., and Howitt, P. (2015). Lessons from Schumpeterian growth theory. The American Economic Review, 105(5), 94-99.

Blind, K., J. Edler, U. Schmoch, B. Anderson, J. Howells, I. Miles, J. Roberts, L. Green, R. Evangelista and Hipp, C. (2003). Patents in the Service Industries – Final Report, Fraunhofer Institut für Systemtechnik und Innovationsforschung, Karlsruhe.

Bresnahan, T. J. and Trajtenberg, M. (1995). General Purpose Technologies: 'Engines of Growth'?. Journal of Econometrics, 65 (1), 83-108.

Danguy, J., de Rassenfosse, G. and van Pottelsberghe de la Potterie, B. (2014). On the origins of the worldwide surge in patenting: An industry perspective on the R&D-patent relationship. Industrial and Corporate Change, 23 (2), 535-572.

Dorner, M., Bender, S., Harhoff, D., Hoisl, K. and Scioch, P. (2014). The MPI-IC-IAB-Inventor Data 2002 (MIID2002). Record-Linkage of Patent Register Data with Labor Market Biography Data of the IAB. FDZ Methodenreport 06/2014. Nuremberg.

Eaton, J. and Kortum, S. (2002). Technology, Geography, and Trade. Econometrica, 70 (5), 1741-1779.


[22] Concordance tables (horizontal and vertical) between various hierarchies of the NACE Rev. 1, NACE Rev. 1.1, NACE Rev. 2 and different groupings of IPC classes in technology areas are available in the supplementary materials.

Eberle, J., Jacobebbinghaus, P., Ludsteck, J. and Witter, J. (2011). Generation of time-consistent industry codes in the face of classification changes. Simple heuristic based on the Establishment History Panel (BHP). FDZ-Methodenreport, 05/2011, Nuremberg.

Everson, R. and Putnam, J. (1988). The Yale-Canada patent flow concordance. Yale University Economic Growth Centre Working Paper.

Englander, S. A., Evenson, R. and Hanazaki, M. (1988). R&D, Innovation, and the total factor productivity slowdown. OECD Economic Studies, 11 (Autumn), 7-42.

European Patent Office (EPO) and Office for Harmonization in the Internal Market (OHIM) (Eds.) (2013). Intellectual property rights intensive industries: contribution to economic performance and employment in the European Union. Industry-Level Analysis Report, September 2013. Munich, Alicante. http://documents.epo.org/projects/babylon/eponet.nsf/0/8E1E34349D4546C3C1257BF300343D8B/$File/ip_intensive_industries_en.pdf, accessed: 2017-07-05.

Frietsch, R. and Schmoch, U. (2010). Transnational patents and international markets. Scientometrics, 82 (1), 185-100.

Gambardella, A., Giuri, P. and Mariani, M. (2005). The Value of European Patents: Evidence from a survey of European Inventors – Final report of the PatVal EU Project. URL: http://www.alfonsogambardella.it/PATVALFinalReport.pdf, accessed: 2016-12-10.

Glitz, A. and Meyersson, E. (2017). Industrial espionage and productivity, IZA Discussion Paper No. 10816.

Greif, S. and Potkowik, G. (1990). Patente und Wirtschaftszweige, Zusammenführung der Internationalen Patentklassifikation und der Systematik der Wirtschaftszweige, Heymann, Köln.

Griliches, Z. (1990). Patent Statistics as Economic Indicators: A Survey. Journal of Economic Literature, 28 (4), 1661-1707.

Johnson, D. K. N. (2002). The OECD Technology Concordance (OTC). Patents by Industry of Manufacture and Sector of Use. OECD STI Working Papers, 2002/5, Paris.

Kortum, S. and Putnam, J. (1997). Assigning Patents to Industries: Tests of the Yale Technology Concordance. Economic Systems Research, 9 (2), 161-176.

Kronz, H. and Grevink, H. (1980). Patent statistics as indicators of technological and commercial trends in the member States of the European Communities (EEC). World Patent Information, 2 (1), 4-12.

Lybbert, T. J. and Zolas, N. J. (2014). Getting patents and economic data to speak to each other: An 'Algorithmic Links with Probabilities' approach for joint analyses of patenting and economic activity. Research Policy, 43 (3), 530-542.

Maurseth, P. B., & Verspagen, B. (2002). Knowledge spillovers in Europe: a patent citations analysis. The Scandinavian journal of economics, 104(4), 531-545.

Meliciani, V. (2000). The relationship between R&D, investment and patents: a panel data analysis. Applied Economics, 32 (11), 1429-1437.

Schild, C. J. (2016). Linking 'Orbis' Company Data with Establishment Data from the German Federal Employment Agency. German Record Linkage Center Working Paper 2016-02. Nuremberg.

Schmoch, U., Laville, F., Patent, P. and Frietsch, R. (2003). Linking Technology Areas to Industrial Sectors – Final Report to the European Commission, November 2003.

Schmoch, U. (2008). Concept of a Technology Classification for Country Comparisons – Final Report to the World Intellectual Property Organisation (WIPO). URL: http://www.wipo.int/export/sites/www/ipstats/en/statistics/patents/pdf/wipo_ipc_technology.pdf, accessed: 2016-02-04.

Schmucker, A., Seth, S., Ludsteck, J. Eberle, J. and Ganzer, A. (2016). Establishment History Panel 1975-2014. FDZ-Datenreport, 03/2016, Nuremberg.

Van Looy, B., Vereyen, C. and Schmoch, U. (2014). Patent Statistics: Concordance IPC V8 – NACE Rev.2. EUROSTAT. Url: https://circabc.europa.eu/sd/a/d1475596-1568-408a-9191-426629047e31/2014-10-16-Final%20IPC_NACE2_2014.pdf, accessed: 2016-09-19.

Verspagen, B., van Moergastel, T. and Slabbers, M. (1994). MERIT concordance table: IPC - ISIC (Rev.2). MERIT Research Memorandum 2/94-0.

**Table 1** Patenting activity by technology area (TF34) in linked inventor-establishment data

| Rank | Technology area (TF34) | Patent-inventor records | | | |
|---|---|---|---|---|---|
| | | count (*inv*) | | fractional count (*frac_inv*) | |
| | | # | % | # | % |
| 1 | 31 Transport | 43,275 | 8.25 | 20,481 | 8.68 |
| 2 | 1 Electrical machinery, apparatus, energy | 34,375 | 6.56 | 16,225 | 6.88 |
| 3 | 30 Mechanical elements | 26,369 | 5.03 | 12,895 | 5.47 |
| 4 | 10 Measurement | 25,474 | 4.86 | 12,381 | 5.25 |
| 5 | 26 Engines, pumps, turbines | 25,437 | 4.85 | 11,678 | 4.95 |
| 6 | 14 Organic fine chemistry | 40,375 | 7.70 | 11,495 | 4.87 |
| 7 | 25 Machine tools | 17,692 | 3.37 | 9,129 | 3.87 |
| 8 | 13 Medical technology | 17256 | 3.29 | 8687 | 3.68 |
| 9 | 28 Other special machines | 17,568 | 3.35 | 8,449 | 3.58 |
| 10 | 34 Civil engineering | 14,173 | 2.70 | 8,344 | 3.54 |
| 11 | 3 Telecommunications | 15,972 | 3.05 | 8,261 | 3.50 |
| 12 | 24 Handling | 14,770 | 2.82 | 8,160 | 3.46 |
| 13 | 6 Computer technology | 14,756 | 2.81 | 7,424 | 3.15 |
| 14 | 22 Chemical engineering | 17,062 | 3.25 | 7,165 | 3.04 |
| 15 | 19 Basic materials chemistry | 21,356 | 4.07 | 7,142 | 3.03 |
| 16 | 27 Textile and paper machines | 14,661 | 2.80 | 6,769 | 2.87 |
| 17 | 17 Macromolecular chemistry, polymers | 18,987 | 3.62 | 6,624 | 2.81 |
| 18 | 15 Biotechnology | 13,903 | 2.65 | 5,750 | 2.44 |
| 19 | 16 Pharmaceuticals | 14,633 | 2.79 | 5,723 | 2.43 |
| 20 | 29 Thermal processes and apparatus | 11,375 | 2.17 | 5,007 | 2.12 |
| 21 | 4 Digital communication | 9,261 | 1.77 | 4,786 | 2.03 |
| 22 | 12 Control | 9,235 | 1.76 | 4,605 | 1.95 |
| 23 | 20 Materials, metallurgy | 11,221 | 2.14 | 4,558 | 1.93 |
| 24 | 21 Surface technology, coating | 10,851 | 2.07 | 4,482 | 1.90 |
| 25 | 33 Other consumer goods | 9,314 | 1.78 | 4,382 | 1.86 |
| 26 | 2 Audio-visual technology | 8,735 | 1.67 | 4,226 | 1.79 |
| 27 | 8 Semiconductors | 9,705 | 1.85 | 4,212 | 1.79 |
| 28 | 9 Optics | 9,070 | 1.73 | 3,949 | 1.67 |
| 29 | 32 Furniture, games | 7,757 | 1.48 | 3,834 | 1.63 |
| 30 | 23 Environmental technology | 7,572 | 1.44 | 3,395 | 1.44 |
| 31 | 5 Basic communication processes | 3,183 | 0.61 | 1,788 | 0.76 |
| 32 | 11 Analysis of biological materials | 4,123 | 0.79 | 1,704 | 0.72 |
| 33 | 18 Food chemistry | 2,762 | 0.53 | 1,285 | 0.54 |
| 34 | 7 IT methods for management | 2,128 | 0.41 | 938 | 0.40 |
| | **Total** | **524,386** | **100** | **235,933** | **100** |

**Table 2** Patenting activity by NACE Rev. 2 divisions (2-digit numerical codes) in linked inventor-establishment data

| Rank | NACE Rev. 2 / WZ 2008 divisions (two-digit numerical codes) | Patent-inventor records | | | |
|---|---|---|---|---|---|
| | | count (*inv*) | | fractional count (*frac_inv*) | |
| | | # | % | # | % |
| 1 | 28 Manufacture of machinery and equipment n.e.c. | 67,867 | 12.94 | 34,340.84 | 14.56 |
| 2 | 26 Manufacture of computer, electronic and optical products | 59,107 | 11.27 | 28,885.69 | 12.24 |
| 3 | 20 Manufacture of chemicals and chemical products | 75,850 | 14.46 | 25,196.87 | 10.68 |
| 4 | 29 Manufacture of motor vehicles, trailers and semi-trailers | 55,567 | 10.60 | 25,136.81 | 10.65 |
| 5 | 72 Scientific research and development | 51,493 | 9.82 | 20,464.74 | 8.67 |
| 6 | 27 Manufacture of electrical equipment | 41,693 | 7.95 | 18,374.31 | 7.79 |
| 7 | 25 Manufacture of fabricated metal products, except machinery and equipment | 17,231 | 3.29 | 9,282.49 | 3.93 |
| 8 | 46 Wholesale trade, except of motor vehicles and motorcycles | 15,087 | 2.88 | 7,893.42 | 3.35 |
| 9 | 71 Architectural and engineering activities; technical testing and analysis | 14,009 | 2.67 | 6,818.97 | 2.89 |
| 10 | 21 Manufacture of basic pharmaceutical products and pharmaceutical preparations | 20,664 | 3.94 | 6,686.72 | 2.83 |
| 11 | 22 Manufacture of rubber and plastic products | 11,075 | 2.11 | 5,728.43 | 2.43 |
| 12 | 32 Other manufacturing | 9,570 | 1.82 | 4,661.85 | 1.98 |
| 13 | 70 Activities of head offices; management consultancy activities | 9,011 | 1.72 | 4,612.04 | 1.95 |
| 14 | 85 Education | 9,216 | 1.76 | 4,576.02 | 1.94 |
| 15 | 62 Computer programming, consultancy and related activities | 7,990 | 1.52 | 4,057.19 | 1.72 |
| | Others industry divisions (ranks 16-86) | 58,956 | 11.24 | 29,216.61 | 12.38 |
| | **Total** | **524,386** | **100** | **235,933** | **100** |

**Table 3** Correlations of weights over time

| | Pooled | Year 2000 | Year 2005 | Year 2010 |
|---|---|---|---|---|
| *Horizontal correspondence table (34 technology areas)* | | | | |
| Pooled | 1.000 | | | |
| Year 2000 | 0.973 | 1.000 | | |
| Year 2005 | 0.981 | 0.930 | 1.000 | |
| Year 2010 | 0.915 | 0.845 | 0.900 | 1.000 |
| *Vertical correspondence table (86 NACE Rev. 2 divisions)* | | | | |
| Pooled | 1.000 | | | |
| Year 2000 | 0.773 | 1.000 | | |
| Year 2005 | 0.741 | 0.536 | 1.000 | |
| Year 2010 | 0.608 | 0.428 | 0.429 | 1.000 |

**Table 4** Comparison with ALP concordance table

| | NACE Rev. 2 to IPC4 (vertical table) | IPC4 to NACE Rev. 2 (horizontal table) |
|---|---|---|
| | (1) | (2) |
| *Ranks* | | |
| Rank 1 = rank 1 in ALP (%) | 53.33 | 50.00 |
| Rank 1 = top3 rank in ALP (%) | 84.28 | 59.57 |
| *Bivariate correlations of weights* | | |
| Weight and weight ALP | 0.373 | 0.630 |
| Rank 1 = rank 1 in ALP | 0.131 | 0.938 |
| Top3 rank = top3 rank in ALP | 0.172 | 0.697 |
| *Differences in the weights* | | |
| Mean [Std. dev.] of weight - weight ALP | -0.031 [0.081] | -0.117 [0.313] |
| IPC4-industry observations | N = 1,775 | N = 518 |

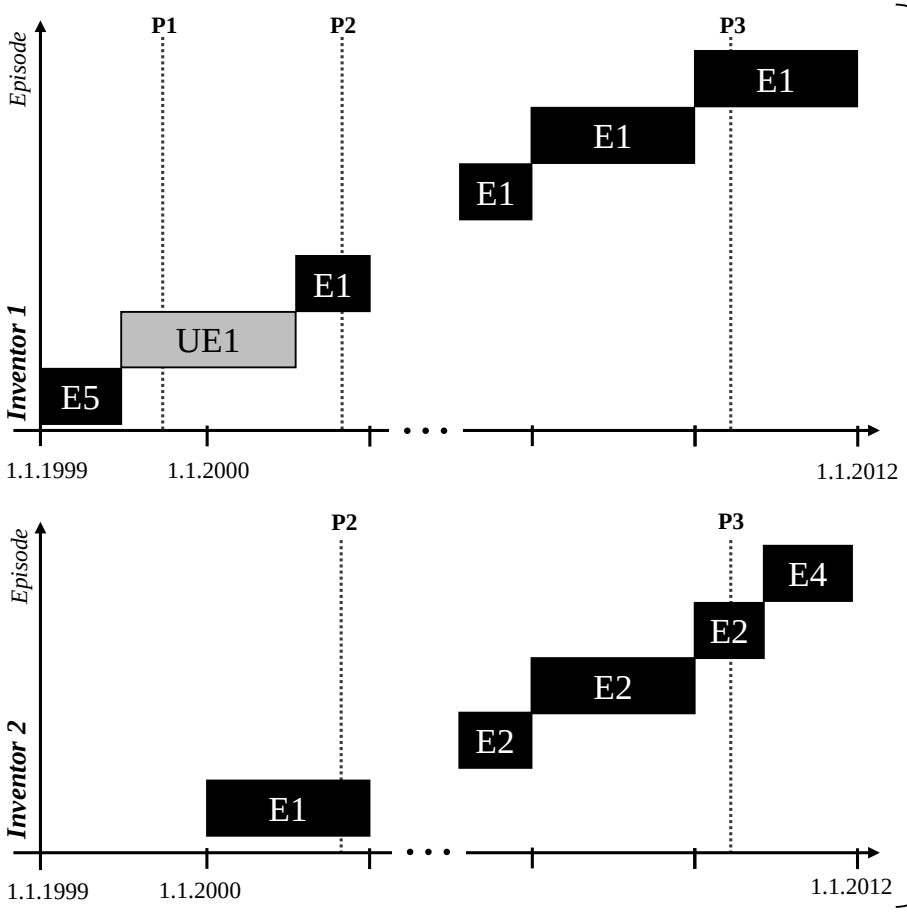
**Table 5** Bivariate correlations of patent intensities and innovation indicators for NACE Rev. 2 industries

| Bivariate correlations<br>Variable and patent intensities (patents / 1,000 employees) at the industry level (NACE Rev. 2) | Data | All divisions (01-99) | | Manufacturing divisions (10-33) | | ZEW-MIP sectors | |
|---|---|---|---|---|---|---|---|
| | | National pat. | Transnat. pat. | National pat. | Transnat. pat. | National pat. | Transnat. pat. |
| | | (1) | (2) | (3) | (4) | (5) | (6) |
| Highly skilled employees (%) | IAB-BHP | 0.415 | 0.408 | **0.766** | **0.753** | 0.511 | 0.515 |
| Highly skilled science & engineering employees (%) | IAB-BHP | **0.565** | **0.612** | **0.654** | **0.749** | **0.708** | **0.677** |
| Firms with new to market product innovations (%) | IAB-BP | 0.410 | 0.414 | **0.621** | **0.598** | - | - |
| Share of innovators (%) | ZEW-MIP | - | - | - | - | **0.654** | **0.650** |
| Share of firms with continuous R&D activities (%) | ZEW-MIP | - | - | - | - | **0.820** | **0.859** |
| Innovation intensity (%) | ZEW-MIP | - | - | - | - | **0.773** | **0.799** |
| Share of firms with new-to-market product innovations (%) | ZEW-MIP | - | - | - | - | **0.759** | **0.759** |
| Avg. share of turnover from new-to-market product innovations (%) | ZEW-MIP | - | - | - | - | **0.737** | **0.654** |
| Number of industry-year observations in sample | | N = 591 | | N = 168 | | N = 258 | |

Notes: Bivariate correlations exceeding the threshold of rho > 0.5 are printed in bold; Data sources: IAB-BHP = Establishment History Panel of IAB; IAB-BP = IAB Establishment Panel Survey; ZEW-MIP = ZEW Mannheim Innovation Panel Survey; Definition of national and transnational patents adopted from Frietsch and Schmoch (2010).

**Figure 1** Data structure of linked inventor-establishment data

**Sample inventor biographies for inventors I1 and I2**




**Representation of patents P1, P2 and P3 in the linked inventor-establishment data**

| appln_id | tech (*t*) | N_inv | inv_id | est_id | nace (*i*) | inv | frac_inv |
|---|---|---|---|---|---|---|---|
| **P1** | **11** | **1** | **I1** | **n.a.** | **n.a.** | n.a. | n.a. |
| **P2** | **24** | **4** | **I1** | **E1** | **25** | 1 | 0.25 |
| **P2** | **24** | **4** | **I2** | **E1** | **25** | 1 | 0.25 |
| P2 | 24 | 4 | I3 | E1 | 25 | 1 | 0.25 |
| P2 | 24 | 4 | I4 | E7 | 72 | 1 | 0.25 |
| **P3** | **22** | **2** | **I2** | **E2** | **12** | 1 | 0.5 |
| **P3** | **22** | **2** | **I1** | **E1** | **25** | 1 | 0.5 |
| | | | | | $y_{it}$ | **7** | **2** |

Variables: *appln_id* = patent id, *tech* = technology class *t* , *N_inv* = total number of inventors, *inv_id* = inventor id, *est_id* = establishment id, *nace* = NACE industry class *i*, *inv* = unique patent-inventor count, *frac_inv* = fractional patent-inventor count (*inv* / *N_inv*).
Note: Matched patent-inventor/employee records printed in bold.

**Legend**

E# — Employment episode of inventor *k* in establishment *E#* in industry *i* (IAB data).

UE# — Unemployment episode of inventor *k* (IAB data).

⋮ — Patent application of inventor *k* listed on patent *p* in technology *t* (Patstat data).

**Figure 2** Industry (NACE Rev. 2 divisions) – technology (TF34) concordance table, horizontal structure

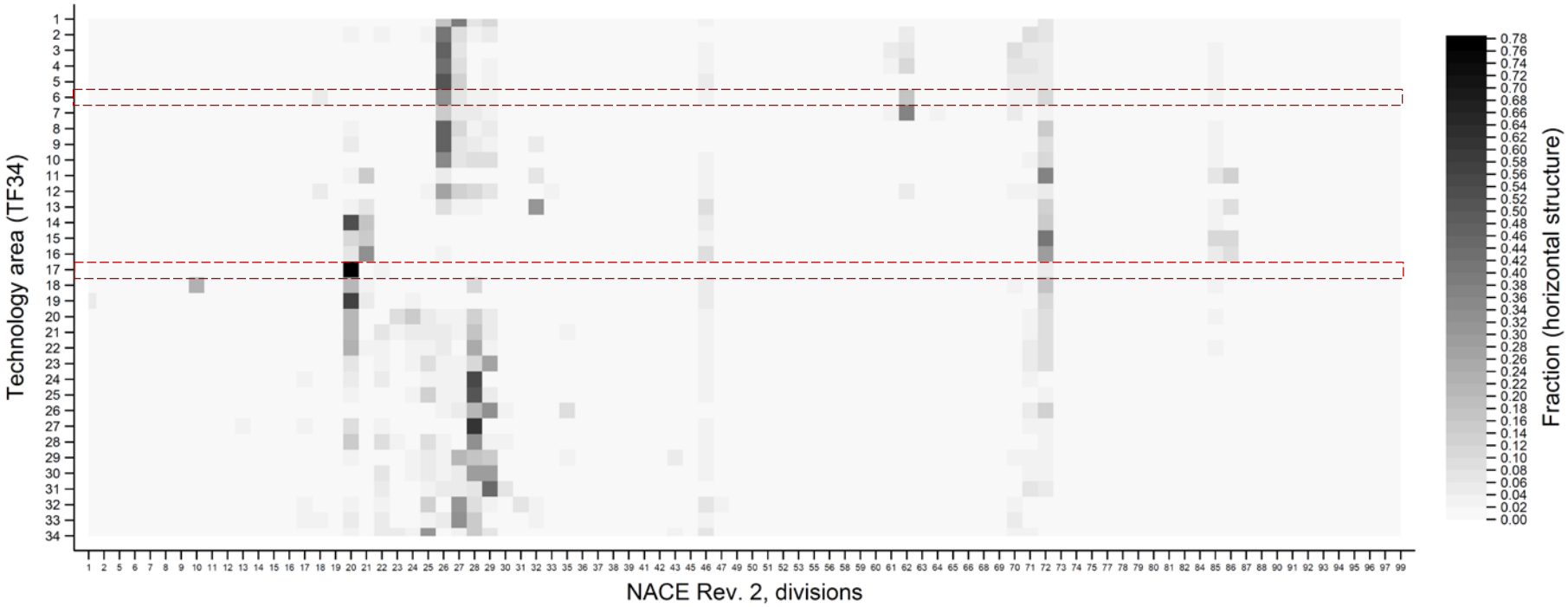


**Figure 3** Industry (NACE Rev. 2 divisions) – technology (TF34) concordance table, vertical structure

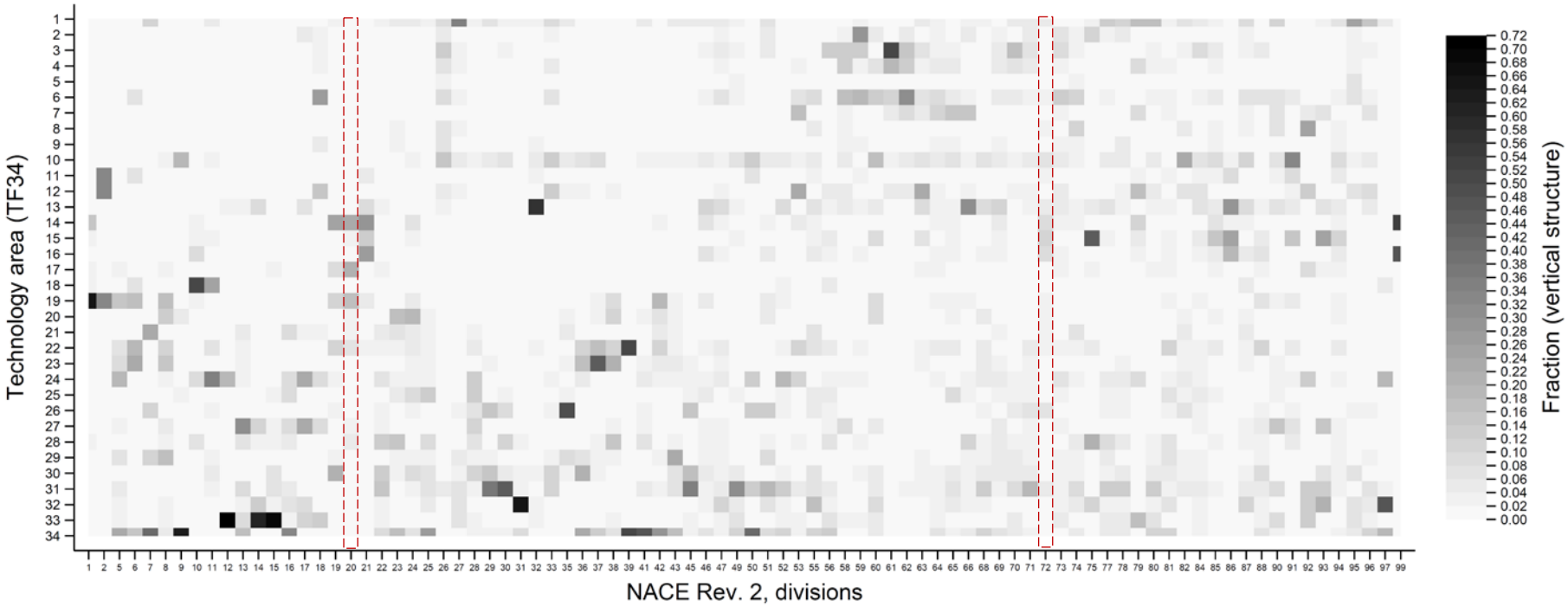
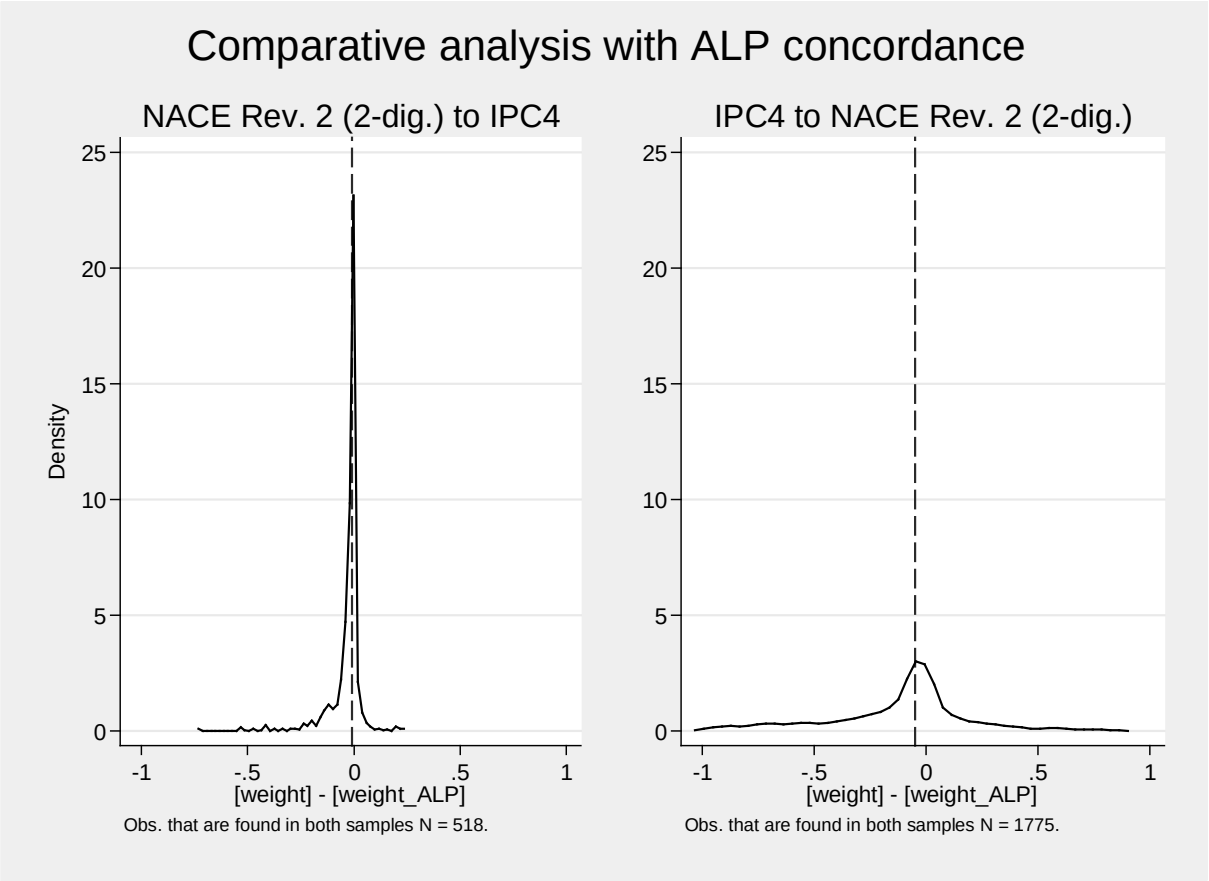


**Figure 4** Differences in the weights compared to ALP concordance



## Appendix

**Table A1** Correlation of technology area (TF34) shares across different patent samples

| | | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| 1 | EP Patents 1999-2011 | 1.000 | | | |
| 2 | Linked EP Patents 1999-2011 | 0.990 | 1.000 | | |
| 3 | National priority patents | 0.924 | 0.923 | 1.000 | |
| 4 | Transnational patents | 0.989 | 0.983 | 0.955 | 1.000 |

Note: Correlations computed over population shares of 34 technology areas in the different patent samples.

**Table A2** Concentration indicators within industry-technology concordance tables

| | Obs. | Mean | Std. Dev. | Min. | Max. |
|---|---|---|---|---|---|
| *Horizontal correspondence table (34 technology areas)* | | | | | |
| HHI (computed over industries) | 34 | 0.213 | 0.103 | 0.098 | 0.621 |
| Modal value of industry | 34 | 0.389 | 0.129 | 0.205 | 0.785 |
| Top 3 ranked industries | 34 | 0.622 | 0.095 | 0.458 | 0.868 |
| *Vertical correspondence table (86 NACE Rev. 2 divisions)* | | | | | |
| HHI (computed over technologies) | 86 | 0.162 | 0.124 | 0.042 | 0.542 |
| Modal value of technology | 86 | 0.278 | 0.160 | 0.082 | 0.710 |
| Top 3 ranked technologies | 86 | 0.523 | 0.199 | 0.217 | 1.000 |

**Table A3** Summary statistics of innovation indicators

| Variables | Data | All divisions (1-99) | | Manufacturing divisions (10-33) | | ZEW-MIP sectors | |
|---|---|---|---|---|---|---|---|
| | | Mean | Std. dev. | Mean | Std. dev. | Mean | Std. dev. |
| National patents (per 1,000 emp.) | PATSTAT | 0.850 | 1.882 | 2.029 | 2.170 | 2.674 | 3.986 |
| Transnational patents (per 1,000 emp.) | PATSTAT | 1.494 | 3.139 | 3.578 | 3.667 | 1.546 | 2.424 |
| Highly skilled employees (per 1,000 emp., %) | IAB-BHP | 13.791 | 9.872 | 11.367 | 6.797 | 15.297 | 10.593 |
| Highly skilled science & engineering employees (per 1,000 emp., %) | IAB-BHP | 15.249 | 15.888 | 24.824 | 15.194 | 3.775 | 5.206 |
| Firms with new-to-market product innovations (%) | IAB-BP | 10.450 | 10.723 | 15.727 | 8.644 | - | - |
| Share of innovators (%) | ZEW-MIP | - | - | - | - | 4.196 | 7.496 |
| Share of firms with continuous R&D activities (%) | ZEW-MIP | - | - | - | - | 18.302 | 18.214 |
| Innovation intensity (%) | ZEW-MIP | - | - | - | - | 50.853 | 18.096 |
| Share of firms with new-to-market product innovations (%) | ZEW-MIP | - | - | - | - | 16.702 | 12.345 |
| Avg. share of turnover from new-to-market product innovations (%) | ZEW-MIP | - | - | - | - | 3.460 | 3.178 |
| Number of industry-year observations in sample | | N = 591 | | N = 168 | | N = 258 | |

Notes: Data sources: IAB-BHP = Establishment History Panel of IAB; IAB-BP = IAB Establishment Panel Survey; ZEW-MIP = ZEW Mannheim Innovation Panel Survey; Definition of national and transnational patents adopted from Frietsch and Schmoch (2010).