The World Economy

The World Economy (2016)
doi: 10.1111/twec.12382

An 'Algorithmic Links with Probabilities' Concordance for Trademarks with an Application Towards Bilateral IP Flows

Nikolas Zolas[1], *Travis J. Lybbert*[2] *and Prantik Bhattacharyya*[3]
[1]*Center for Economic Studies, United States Census Bureau, Washington, DC, USA,* [2]*Agricultural & Resource Economics, University of California – Davis, Davis, CA, USA, and* [3]*Computer Science, University of California – Davis, Davis, CA, USA*

1. INTRODUCTION

IN the contemporary global economy, trademarks (TMs) play an important role in a wide array of industries and sectors and shape the competitive landscape of many diverse markets. Although reliance on TMs certainly evolves with structural changes and economic development, the economic importance of TMs is as apparent in developed countries as it is in emerging and even developing countries. Despite these realities, economists and policy analysts alike have been unable to conduct careful empirical analysis of TMs in the modern economy because TM data and economic activity data are organised differently and can therefore not be analysed jointly. In this project, we aim to remedy this incompatibility by building a bridge between TM and economic data that enables these data to 'speak to each other'.

This work is motivated by the paucity of rigorous empirical research into the relationship between TMs and economic activity, which is due to two primary facts. First, the competitive and strategic considerations that shape whether and how firms rely on TMs to build brands and differentiate their products and services differ dramatically across industrial sectors. This implies that any empirical analysis should either focus on TM activity in a particular sector or otherwise allow for substantially different empirical relationships between TMs and economic activity across sectors. Second, while TM data are available from more and more countries and economic data are widely available at a high resolution of industrial sector or product category, merging these data by linking the goods & services (GS) covered by a TM to sectors or products is difficult and – to date – has been very limited. This presents a serious constraint on matching TM and economic data at a useful level of resolution and in a robust and reliable way. In conjunction with the first fact, this severely limits the kinds of empirical research that are possible in this area.

We develop an algorithmic approach we call 'algorithmic links with probabilities' (ALP) matching to explicitly link TM and economic data via standard, widely used product and industry classification systems such as the Standard International Trade Classification (SITC), International Standardized Industrial Classification (ISIC), Harmonized System (HS), and North

Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the US Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed. We thank members of the US Patent and Trademark Office (USPTO), World Intellectual Property Organization (WIPO) and participants of the USPTO-NYU Trademark Workshop for comments and assistance with the data. Portions of this paper were originally prepared as a report for WIPO with support from WIPO.

American Industry Classification System (NAICS). As a key benefit to this approach, these class-level ALP concordances implicitly reflect differences in TM usage across economic sectors – and therefore link TMs to economic activity according to predominant TM-use patterns.

This ALP matching approach, which has been used similarly to concord patents to economic data (Lybbert and Zolas, 2014), enables researchers to map TM data directly into trade or industry categories in order to create measures of TM-use intensity that are comparable across countries and over time and to empirically model the determinants of international TM flows and the economic effects of TMs. Together with similar ALP concordances designed for mapping patents into the same economic classification systems, these new data tools open up broader possibilities to jointly analyse TMs and patents. Given how much intellectual property strategies vary from industry to industry and given the interdependence that is often evident in the use of these two important forms of intellectual property, the ability to combine patent, TM and economic data by industry into a single analysis is particularly potent. Such joint analysis would reflect the inherent heterogeneity in TM usage across sectors described above and would ultimately improve our understanding of the relationships between intellectual property and the value of production of both goods and services domestically and the value of goods traded internationally. Analyses such as these could not only improve our ability to model and understand how TMs fit into the contemporary global economy generally, but would also serve as a platform for addressing a host of policy-relevant research questions.

2. BACKGROUND

Trademark filings have expanded rapidly in recent decades. As described by the 2012 World Intellectual Property Indicators report (WIPO 2012), total TM applications worldwide more than doubled between 1995 and 2011, with more than 4.2 million applications filed in 2011. Much of this growth was driven by TMs filed in and by emerging economies, with China accounting for nearly half of the overall growth between 2004 and 2011 (46.9 per cent). What is somewhat surprising about this growth is that while overall TM output has increased dramatically, the level of foreign TMs (i.e. those registered outside the registrant's home jurisdiction), has more or less stayed flat over this same time period, despite the dramatic increase in trade and other forms of transferred intellectual property, such as patents. Institutional innovations have facilitated these internationally filed TMs. Specifically, the Madrid Protocol became operational in 1996, making it much easier for TM owners to apply for international registrations in countries that have joined this protocol.[1]

Using data from the WIPO, we provide additional perspectives on these trends. We focus mainly on foreign TM registrations (so-called exported TMs) and consider how these exported TMs flowed from and to different country incomes classes during the past two decades. We classify countries into income classes using the World Bank high, middle and low-income categories. Table 1 shows average annual TM registrations sent to and from these different income categories over the years 1994–2011. While high-income countries filed on average 10 times more TM registrations than middle-income countries and more than 100 times more

[1] This protocol materialised from the original Madrid Agreement, which first entered into force in 1892 as a means for international TM registrations and had 56 member countries at the time the protocol was agreed upon. Today, there are 90 member countries in the Madrid Protocol, allowing TM holders to extend the jurisdiction of their TM to anyone of these countries at any time during the life of the TM.

TABLE 1
Average (2004–11) Exported Trademark (TM) Registrations and TM Intensity from and to Different Country Income Classes

Average Exported TM Registrations	*From*						*Total*
	High	*% Total*	*Middle*	*% Total*	*Low*	*% Total*	
To							
High	517,664	58	64,158	48	5,679	45	587,501
Middle	254,932	29	41,481	31	4,571	36	300,983
Low	115,785	13	28,022	21	2,477	19	146,283
Total	888,381		133,660		12,726		

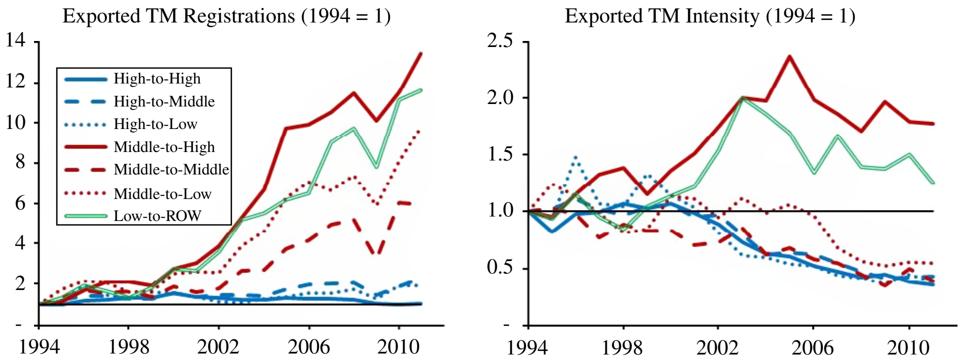
Average Exported TM Intensity	*From*						*Weighted Average*
	High	*% Own*	*Middle*	*% Own*	*Low*	*% Own*	
To							
High	187	100	54	51	21	58	139
Middle	282	151	106	100	40	111	213
Low	388	207	147	139	36	100	262
Weighted average	224		76		28		

than low-income countries, the receipt of these registrations is more equally shared across these income groups.

To enable more direct comparisons of these differences in foreign TM filings, we normalise them by the total value of trade flowing between these income groups. The resulting measure shown in Table 1 – the exported TM intensity – represents the number of TM registrations filed abroad by the countries in a given income category for every \$1 billion of exports from these same countries. While high-income countries register foreign TMs more intensively than middle and low-income countries, middle and low-income countries attract nearly two times more registrations per \$1 billion of exports. The pattern of TM-use intensity from high-income countries is quite distinct: high-income countries on average registered 187 foreign TMs in other high-income countries but roughly 282–388 in middle and low-income countries for every \$1 billion of exports.

It is also informative to see how these TM measures have evolved overtime. Figure 1 shows this evolution since 1994. Since total annual TM registrations from low-income countries are relatively low and volatile, we consider total TMs from low-income countries to the Rest of the World (ROW) instead of by income category. Considering the exported TMs first (left), we see a dramatic expansion of foreign TM registrations filed by middle and low-income countries. Filings from middle-to-high-income countries have increased nearly 14 times during this period. The impact of the economic downturn in 2009 appears to have been short-lived as registrations continue to grow. Based on TM intensity measures (right panel), only low-to-ROW and middle-to-high TM registrations have grown faster than exports. While most TM intensity rates have steadily declined by half, the intensity of TM use from middle-to-high-income countries has nearly doubled. This seems to suggest that TMs play a particularly important strategic role for firms in middle and low-income countries looking to expand into higher value markets. In this sense, TMs may enable the integration of such firms into global value chains.

FIGURE 1
Total Exported Trademark (TM) Registrations (Left) and TM Intensity for Exported Registrations (Right) from and to Different Income Classes (ROW = Rest of World)



a. Empirical Research in the Economics of Trademarks

Trademarks are used to differentiate between goods and services offered by competitors within a particular industry. The TM is intended to reveal information to the consumer regarding both the quality and consistency of a line of goods and services (Economides, 1987; Landes and Posner, 1987). For TM holders, TMs provide the ability to bypass retailers and communicate directly with customers, along with the flexibility to expand into other product lines and license the TM to third parties. The economic interpretation of TMs and why they are important stem from the inherent value in promoting market efficiency and market power while reducing rent-seeking behaviour (Ramello, 2006), information asymmetry (Economides, 1987) and search costs (Landes and Posner, 1987). These intangible components make it difficult to assign economic values to TMs; thus, the scope of TM use in economic studies has been somewhat limited.

The use of TMs and how it allows the firm to establish and build a particular brand has been rigorously studied within business under ‘brand management’. Schautschick and Greenhalgh (2013) provide a helpful and thorough overview of existing studies that utilise TM data. In economic studies, TMs have most widely been used in micro-level studies as a proxy for innovation (Schmoch, 2003; Mendonca et al., 2004; Malmberg, 2005; Millot, 2009; Greenhalgh and Rogers, 2007), but also in distinguishing the usage of TMs across firm size (Allegrazza and Guard-Rauchs, 1999; Greenhalgh et al., 2001; Mainwaring et al., 2004) and industry (Scherer, 1983; Greenhalgh et al., 2001; Loundes and Rogers, 2003; Schmoch, 2003; Jensen and Webster, 2004; Mainwaring et al., 2004). These findings suggest that TMs serve as reasonable proxies for innovation in certain industries, like pharmaceuticals, and less well for others such as the electromechanical and automotive industries (Malmberg, 2005). In Mendonca et al. (2004), the authors suggest several ways in which TMs can be used to analyse certain relevant aspects of innovation and industrial change. They encourage greater studies that use TM data and explain how TM-based indicators can provide a partial measure of innovative firm output, international patterns of specialisation, links between technology and marketing, as well as the evolution of firm organisation and structure.

There are a limited number of papers using aggregate measures of TMs to economic outcomes. Yorukoglu (2000) uses TMs to infer economic growth, while Fink et al. (2003) and

Mangani (2007) use TMs to measure trade specialisation. Outside of these limited studies, the use of TMs in economic studies has been limited because of the difficulty in assigning economic values to TMs and problems with aggregation since TMs for the same product line can be applied for across multiple goods and services. A proper concordance will be able to address both issues since we will then be able to assign other measurable economic indicators to TM activity, as well as decompose the use of TMs across many different sectors. Economists will better understand how TMs fit into the overall innovation chain, as well as estimate the value of TMs from their different uses in a variety of industries. Matching TM data with trade flows will also provide information regarding the exporting behaviour of firms, quality within and across varieties and intellectual property rights, since TMs can lengthen the period of protection once patents have expired (Rujas, 1999).

b. Key Challenges to Linking TMs to Economic Data

Before describing the approach we have developed, it is important to appreciate the challenges inherent in linking TM data to economic data. The TM system uses the NICE classification scheme. The standard industrial and trade classification schemes are the ISIC system, SITC system, HS and NAICS, respectively. A conventional concordance approach would link a classification level in NICE to a comparable classification level in ISIC, SITC, HS or NAICS.

Unfortunately, such a conventional approach is complicated by the fact that the NICE system is structured very differently than these industry classifications. Each industry classification system is designed to facilitate the collection and processing of data and therefore have an explicit multilevelled hierarchical structure. The NICE scheme, on the other hand, is designed to facilitate the registration of TMs and the subsequent protection of their legal scope – and lacks a comparable hierarchical structure. Although the complete NICE system includes 'basic numbers' for thousands of predefined GS indicators within each of 45 classes, these numbers are used to compare different translated versions of NICE rather than to reference TM applicants' selection of GS indicators that pertain to their TM. Furthermore, in most jurisdictions (including both the US Patent and Trademark Office (USPTO) and the Madrid System) many or even most TMs are registered with user-defined GS indicators (i.e. applicants write their own indicator rather than choosing from those proposed by NICE), which do not explicitly link to indicators with 'basic numbers' in NICE. As a result of how the NICE classification scheme is used in practice, TM data are typically only explicitly structured according to broad NICE classes and not to the much more specific GS indicators.

Two challenges emerge from this mismatch between the NICE classification system and economic classification systems. A third challenge emerges from how the TM system is used. First, although it would be most useful for many empirical analyses to match TMs to economic activity data at the GS indicator level, it is impossible to do so with a conventional approach because TM data are not organised by GS indicators. One possible remedy to this problem would be to directly classify each TM registration according to industry class. This approach would generate a supplementary data file for any given TM database that contains a list of industry codes at an appropriate level of resolution (e.g. 4-digit) with which each TM in the database is associated – potentially along with a probability that indicates the likelihood (or strength) of the linkage. While this may be technically feasible, the third challenge we describe below at least partially limits the appeal of this approach.

Second, for some policy analyses matching TMs to economic activity at the NICE class level may be genuinely useful, but manually constructing a class-level concordance is challenging because, at a broad level, the NICE scheme is structured differently than SITC, ISIC, HS or NAICS – and consequently each NICE class potentially maps to multiple industry categories and vice versa. In the face of one-to-many matches, it is unclear how to manually determine the weights to use for these multiple matches.

A third challenge to linking TMs to industry is related to how the TM system is used, rather than to the structure of the NICE scheme *per se*. In many jurisdictions, TM applicants can defensively select multiple GS indicators in multiple classes[2] – even if they never intend to use the TM for all of the selected GSs. Ideally, we would link a TM to economic data that is relevant to how the TM is actually used. Defensive selection of GSs implies that identifying which GS indicators are most relevant will be difficult and, if it is possible at all, will require additional effort. The only major jurisdiction that requires TM applicants to subsequently file a specimen that justifies the claimed GS on a given TM registration is the US. In most other jurisdictions, a fee charged per claimed class provides an incentive for applicants not to claim many GSs in several different classes. Although this discourages TM owners from claiming several classes, such a fee structure does nothing to curb defensive claiming within a class once the applicant has decided to claim (and pay for) a given class. One upshot of this important difference in how TM systems work in practice is that directly classifying each TM by industry is likely to be noisier outside than inside the US.

3. 'ALGORITHMIC LINKS WITH PROBABILITIES' MATCHING FOR TRADEMARKS

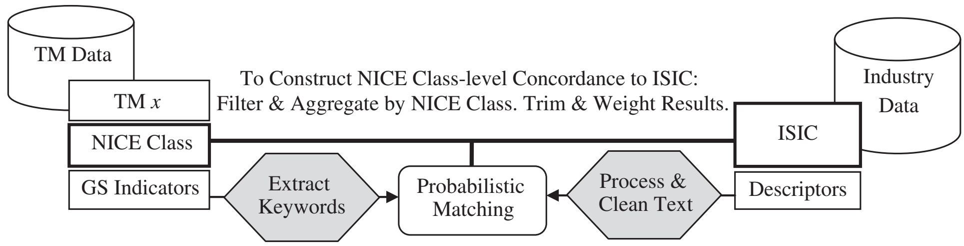
With these challenges in mind, we propose an algorithmic approach that uses data mining and matching as the basis for mapping TM data to economic activity and vice versa. Because the approach relies on computer search algorithms to construct probabilities that indicate the likelihood of a linkage, we call the approach ALP matching. This approach has been used elsewhere to match patent data to economic activity (Lybbert and Zolas, 2014), but applying ALP matching to TMs has required some modifications. Specifically, because TMs lack the textual richness of patents, the basis for matching is more constrained in the case of TMs. This necessitates a different matching approach.

Algorithmic links with probabilities matching is based on linking an individual TM (e.g. TM x) to the categories of an economic classification system (e.g. four-digit ISIC categories). This is done by matching keywords and phrases from the GS indicators for a given TM with the descriptors for each of the economic classification categories. The matches are then reweighted to minimise type I and type II errors.

The database that forms the basis of the concordance is the USPTO TM registrations available via Google.[3] While these data are available from 1884 to the present day, we focus our mapping on the most recent years only, processing the 3,293,150 TMs registered since 1990. Although the ALP methodology we devise can technically be applied to any TM data, the aforementioned fact that the USPTO requires applicants to show proof of use of the TM that

[2] As a related feature of TM policy, some jurisdictions (e.g. China) allow only one class to be designated on each TM application, which means that defensively indicating GSs across multiple classes requires the applicant to submit multiple TM applications. Our work in this project will have to take this into account, but this is a less troublesome problem in many respects.
[3] http://www.google.com/googlebooks/uspto-trademarks.html.

FIGURE 2
Heuristic Depiction of Algorithmic Links with Probabilities Matching Process for Constructing NICE Class-level Concordances to ISIC



conforms to their claimed GS coverage is intended to eliminate defensive GS claims that would introduce noise into the matching process.[4]

For each TM, this database includes – among other things – a description of the TM (including the TM text if it is textual), applicant name, NICE class, and GS indicators. While all of this information is potentially useful for matching a TM to an economic classification category, the GS indicators and the corresponding NICE class are the most useful source of information and are our primary focus. To exploit these GS indicators, we process each TM (e.g. TM *x*) separately and extract keywords and phrases from its listed GS indicators. The full process is generalised in Figure 2, while the following section goes through our methodology in greater detail.

a. Matching

Prior to matching, both the industry descriptions and each TM require extracting and formulating the keywords. For industries, these will be found in the associated descriptive texts that accompany each new revision of the industry classification systems. Often, these descriptions will contain product types, uses and one or two sentences with a brief description of the industry. For our purposes, we utilise the hierarchical structure of each industry classification system and utilise the most disaggregate descriptor available, which will consist of either 4-digit ISIC, 4 to 5-digit SITC, 6-digit HS or 6-digit NAICS codes. We then aggregate in the weighting process to the 2-digit ISIC, 2-digit SITC, 2-digit HS and 4-digit NAICS to ensure that the TM category counts (45) match relatively close to the industry category counts (~100 per industry).[5]

[4] We have, for example, applied this same methodology to the ROMARIN database of Madrid System TM registrations compiled by the World Intellectual Property Organization. In contrast to the domestic TM registrations available in the USPTO database, the ROMARIN database – by definition – includes international registrations exclusively. We do not report the ALP concordances based on this TM database because it is conceptually less appealing due to the frequency of defensive GS claims and reduced number of registrations to search and compile results. A description of the comparison between the ALP concordance constructed using USPTO data versus ROMARIN data is given in the Appendix, where we conclude that differences between concordances are minor.

[5] While technically possible to match to the more disaggregate industry codes, the probability weights turn out to be too small to be of practical use. However, we make these more disaggregate weights available for researchers.

(i) Generating Keywords for Industry Descriptors

Generating the keywords used in the matching process requires several steps. We first take the full text of the description for each corresponding industry and remove generic words that could possibly introduce noise, such as ‘part’, ‘manufacture’ and ‘product’. We also remove the filler words, such as ‘the’ and ‘as’, so that the remaining terms are the most specific and relevant keywords found in the descriptor. In cases where the industry description contained too few keywords or too specific keywords (for instance, some chemical name), then we would augment the keywords using the ‘cross lingual expander’ tool in PATENTSCOPE, a synonym generator specialised to formulate synonyms of words found in patents and other forms of intellectual property.[6] In addition, we augment our keywords with a set of ‘not’ terms, which specify words and constraints that we do not want matched. For instance, when matching the word, ‘sweetened’, we would also pick up ‘unsweetened’, meaning that we would need to include ‘unsweetened’ as a ‘not’ term. Once this process is complete for industries, each disaggregate industry code will have between one and dozens of keywords associated with it. These are then queried and matched directly with the TM keywords.

(ii) Generating Keywords for TMs

To generate keywords for each TM, the process is much different than for industries. The reason being is that whereas for industries, we have several hundred different descriptors, allowing for periodic manual adjustments such as the ‘not’ terms, it would be impossible to comb through the more than 3 million TMs and make any type of manual adjustment outside of pure text extraction. For each TM, we experimented with several different algorithms for extracting keywords from GS indicators. We have also experimented with various ways of expanding these keywords through synonym and other ways.[7] After comparing all these options, we settled on a relatively simple approach that converts each GS indicator phrase – whether predefined or user-defined – into a batch of keywords and expands these keyword batches to include their plural / singular analogs.

For each of the TMs in our data, we extract the text associated with the GS indicator(s). Multiple indicators are separated by semicolon (;). To increase matches, we inspect the density of phrases in the indicator text and process them further to obtain an accurate list of indicator keywords. For many phrases (more than 70 per cent), we run the text through a text-to-keyword extraction software (Topia TermExtract 1.1; https://pypi.python.org/pypi/topia.termextract) and extract keywords. This step preserves the indicator texts supplied by applicants where entries were presumably accurate as keyword-level descriptions (i.e. when the density of key phrases was less than 30 per cent) The end result of this process is a rich set of TM keywords in batches that correspond to the semicolon-delimited GS indicators as chosen by the applicant.

[6] The latest version of PATENTSCOPE can be found here: http://www.wipo.int/PATENTSCOPE/search/clir/clir.jsp?interfaceLanguage=en. We thank Christophe Mazenc at WIPO for his assistance with this step.

[7] For instance, we looked at using company names to extract additional information via name matching with databases of companies that list the industry in which they compete and via Wikipedia entries associated with the company name. We have also experimented with using the TM text to extract additional keywords using Ebay. Although both of these techniques are potentially promising for a subset of TMs, they are ineffective for most TMs.

(iii) Matching

Once the keywords for both the TMs and industries have been generated, the matching process is straightforward. We simply query the full TM database for the keywords generated for each industry and utilise 'batch' matching (i.e. text matches each other perfectly). We retrieve all of the corresponding matching TMs and then pool the TMs by NICE class to generate a frequency of TMs for each industry code. We do not require 100 per cent of the TMs to match to an industry because we have so few NICE classes (45). Instead, we rely heavily on the law of large numbers to provide us with a frequency that is indicative of the true nature between each NICE class and industry. The next section describes additional trimming and reweighting to reduce the potential for type I and type II errors.

b Filtering

The raw matching results potentially map 4, 5 and 6-digit industry codes (of which there exist several hundred for each industry type and revision) with 45 NICE classes. Due to the imbalance between the number of potential industries and number of different NICE classes, early results showed that each NICE class mapped into hundreds of seemingly different and unrelated industries, with corresponding low weights assigned to each mapping. To reduce the imbalance between the number of industries and NICE classes, we employed a 2-digit-level targeted industry filter that excludes non-sensical matches (e.g. a TM claiming only GSs in NICE Class 5 for 'pharmaceutical and veterinary preparations' cannot map to SITC 67 'iron and steel'). We made sure to construct these filters to be generous based on the formal definitions between the NICE class and 2-digit industry descriptors, while simultaneously using a cut-off threshold for aggregated weights (2 per cent). For instance, when a filter case was questionable, we looked at the aggregate weight for the 2-digit industry, and if the frequency was above 2 per cent, then we allowed the 4 and 5-digit industries to map to the corresponding NICE code.

In general, this allows each NICE class to map into targeted industries. This manual process is similar to the one undertaken by Fink et al. (2003) who perform a one-to-one matching of NICE codes to aggregated ISIC codes. However, in our set-up, we allow for more than one match to occur and provide matches at a low level of resolution (2-digit ISIC, SITC, HS or 4-digit NAICS). The final result is that each NICE code has the potential to match up with between 50–100 industry codes on average. This provides us with cleaner frequencies and minimises the potential for type I errors due to certain industries being larger or containing more commonly used words than others.

c. Reweighting

Once the TMs and industries have been matched and filtered, we are left with each NICE class mapping to anywhere between one and dozens of different industries, and vice versa. Next, to further reduce potential errors and/or biases introduced by the matching process, we reweight the results according to the weighting scheme utilised in Lybbert and Zolas (2014). To be specific, we incorporate the 'hybrid' weighting scheme, which was the preferred weighting scheme used in the paper. This weighting scheme is based primarily on Bayes rules, with two adjustments made to account for the fact that some industries and TMs have a greater/lesser propensity to be matched due to the frequency of that class of TM or the broad/specific definition of the TM or industry.

To better illustrate the hybrid weighting scheme, we let A_j be the *ex ante* probability of an industry being matched. with TM class j and B_i be the *ex ante* probability of a TM being matched with industry class i. Assuming J is the total number of different TM classes available (45), Bayes rule gives the *ex post* probability of A_j conditional on observing B_i where:

$$\Pr\,(A_j|B_i) = \frac{\Pr\big(B_i|A_j\big)\Pr\big(A_j\big)}{\Pr(B_i|A_1)\,\Pr(A_1) + \ldots + \Pr(B_i|A_J)\,\Pr(A_J)}. \tag{1}$$

From this, we make two key adjustments. The first adjustment we make gives each industry an equal *ex ante* probability of being matched with TM j (i.e. we set $\Pr(A_j) = 1/j$) so that specifically defined industries/TMs are not penalised and broadly defined industries/TMs are not rewarded. We then counteract the effect of rewarding narrowly defined industries/TMs by again reweighing through the number of actual raw matches ($\Pr(A_j|B_i)$) so that the hybrid weight formula is defined as:

$$W_{ij}^{H} = \frac{\Pr\big(B_i|A_j\big)\big(\Pr\,(A_j|B_i)/J\big)}{\Pr(B_i|A_1)\big(\Pr\,(A_j|B_i)/J\big) + \ldots + \Pr(B_i|A_J)\big(\Pr\,(A_j|B_i)/J\big)}. \tag{2}$$

As Lybbert and Zolas (2014) note, this weighting scheme prioritises (i.e. gives higher weights) to the most frequent matches for narrowly defined industries/TMs, while giving less weight to the broadly defined industries/TMs who might have a large number of erroneous matches due to the nature of their definition.

As a final measure, we impose an additional cut-off condition to remove some of the smaller weights. Our initial cut-off condition was 2 per cent, meaning that matches that had weights below 2 per cent were assigned a weight of zero, and the remaining results were renormalised. This again helps with removing erroneous matches.

Once the full process has been completed, we find that each NICE class maps onto roughly 7–8 industry categories on average, while each industry maps onto approximately 3–4 NICE classes on average. This completes the construction of the concordance. The final results of the concordance for the industry groups ISIC (Rev. 2, 3, 3.1 and 4), SITC (Rev. 2–4), HS (v2002 and 2007) and NAICS (v1997–2007) are posted online.

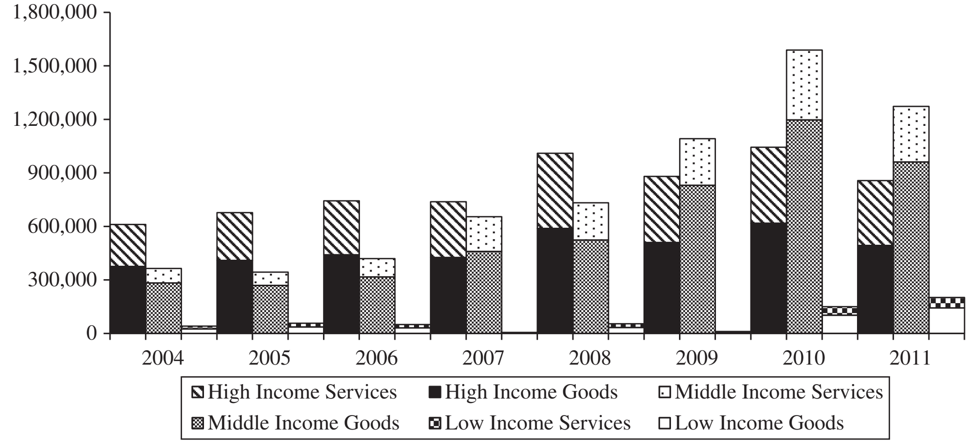
4. USING ALP CONCORDANCES TO JOINTLY ANALYSE TMS AND ECONOMIC DATA

With the concordance, it is possible to jointly analyse industry-level economic activity with TMs, which will allow researchers to address a variety of important, policy-relevant questions more rigorously and at higher resolution and to better understand the value of trademarking and branding. In this section, we take a step in this direction to illustrate the kinds of analysis enabled by these concordances.

a. Trademark Intensity by Income Group

As a first exercise, we analyse country-level differences in trademarking by industry. Specifically, we look at the trademarking intensity by industry across countries of different income levels to identify patterns of TM growth and specialisation. For data, we use the WIPO IP Statistics website, which contains TM output by NICE class for 192 countries between 2004 and 2011. A summary chart in Figure 3 shows the growth of TMs in Goods (NICE classes 1 through 34) and Services (NICE classes 35 through 45) for three separate

FIGURE 3
Trademark Output in Goods (NICE classes 1–34) and Services (NICE classes 35–45) by Income Group, 2004–11



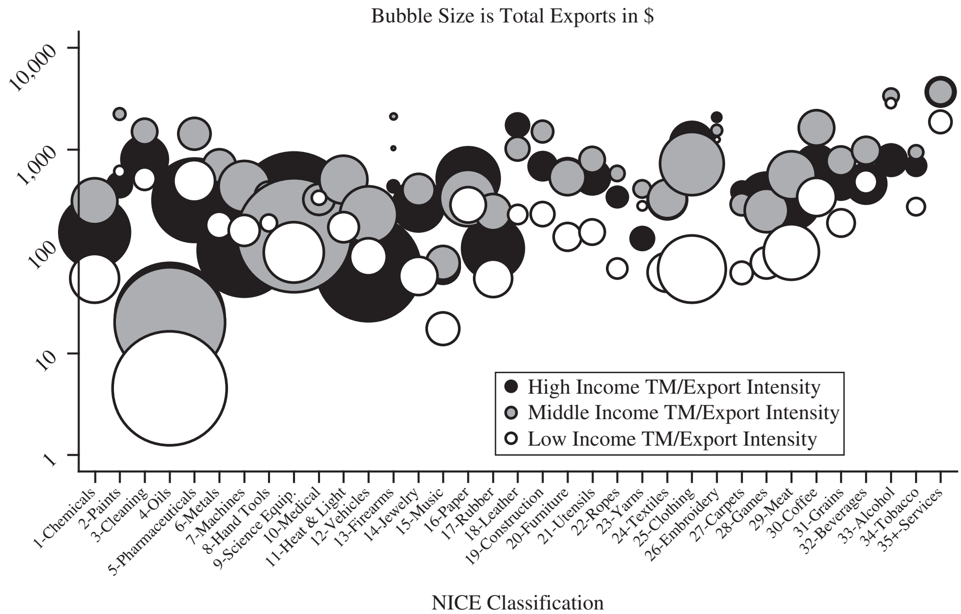
income groups (as defined by the World Bank[8]). We can see that both high-income and middle-income countries experienced rapid growth over this seven-year window with TM output in both goods and services increasing by around 50 per cent in high-income countries. Middle-income countries experienced even more rapid growth and bypassed high-income countries in total TM output, with most of the growth concentrated in Goods. Meanwhile, low-income countries only experienced growth in the most recent years of 2010 and 2011.

To derive our measure of intensity, we proxy for productive output using total export value, since other industry-level data are unavailable for this many countries.[9] The export data are initially organised by two-digit SITC Rev. 2 and was gathered from UN COMTRADE database for the years 2004–11. We applied the ALP Concordance to convert the SITC classification system to the NICE classification and since we are using trade data, focused on trademarked 'Goods' (NICE Class 1 through 34). Figure 4 highlights some interesting patterns in the data.

One of the patterns evident in this figure is that middle-income countries are the most TM-intensive in most TM classes. Averaging across all goods, they use TMs more than twice as intensively as high-income countries. In terms of specific classes, middle and high-income countries are equally intensive in 'clothing, footwear and headgear' (Class 25) and 'leather goods' (Class 18). Meanwhile, middle-income countries are more than ten times as intensive in trademarking in 'pharmaceuticals' (Class 5) and 'yarns and threads' (Class 23). While much of the high intensity of middle-income countries can be attributed to China (for instance, China applied for more than 150,000 TMs in pharmaceuticals between 2004 and 2008 compared to roughly 50,000 TMs applied for by US firms in the same time period), countries such as Russia and Mexico are also very active with trademarking. Low-income countries are most active in 'yarns and threads' (Class 23) and 'alcoholic beverages' (Class 33), with more than four times the intensity of high-income countries in these classes.

[8] Note that the World Bank classifies countries according to four income groups: high, high middle, low middle and low. We combined low- and low middle-income countries to form one low-income group.
[9] We do have industry-level value added and production for OECD countries, which we look at later in the paper.

FIGURE 4
Trademark Intensity (per $ million in Exports) for All Goods by Income Group, 2004–11



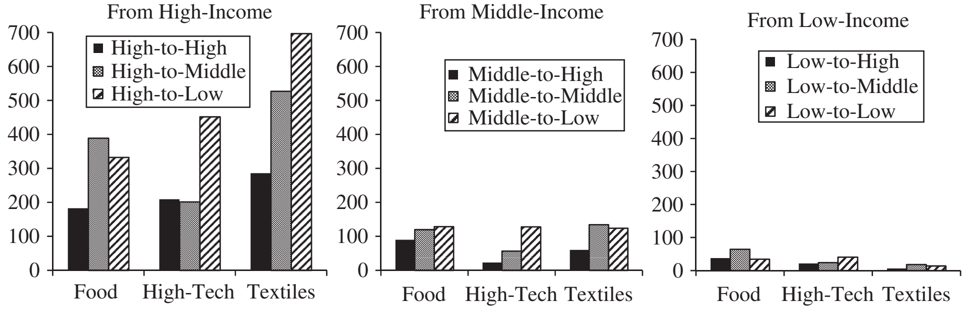
b. Intensity of Foreign Trademark Transfers by Income Group

Next, we look at the intensity of TM flows between income groups for different TM classes, which is an industry-disaggregation of the country-level analysis in Table 1 and Figure 1. We again use total export value as the denominator for relative intensity and focus our attention on a few key TM classes. Specifically, we aggregate NICE classes 29, 30 and 31 to form a broad-level 'food' class. We do the same with classes 23, 24, 25 and 26 to form a 'textile' class. Finally, we combine classes 9, 38 and 42 to form a 'high-tech' goods class. Figure 5 shows the intensity of foreign trademarking relative to exports between each income class.

We can see that foreign TM intensity is highest when a high-income country is the origin. What is more interesting is that the intensity is often highest for transfers to low-income nations (especially for 'high-tech' products) and lowest for transfers to high-income nations. Conducting the same exercise for foreign patents[10] using the ALP Patent Concordance (see Lybbert and Zolas, 2014) yields a very different pattern of intensity: patenting intensity is highest to high-income nations and virtually non-existent to low-income nations (see Figure 6). Clearly, firms incorporate very different strategies for exploiting intellectual property when operating abroad. We speculate that for new technologies that are patentable firms

[10] We use foreign patent data from WIPO's IP Statistics for the same years (2004–08). The patents are initially classified using the International Patent Classification (IPC) system, which differs from the NICE classification system. We first utilise the ALP Patent Concordance to map the patents into SITC Rev. 2. We then layer the ALP Trademark Concordance to map the patents from the SITC into NICE classification system.

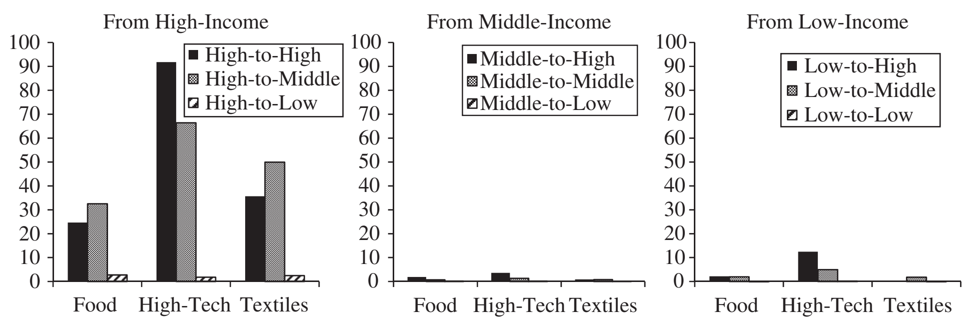
FIGURE 5
Foreign Trademark Intensity (per $ billion in Exports) for Various Goods and Services by Income Group, 2004–11



Note:
'Food' consists of NICE classes 29, 30 and 31. 'High-tech' consists of NICE classes 9 and 42. 'Textiles' consists of NICE classes 23, 24 and 25.

prioritise maintaining their competitive edge *vis-à-vis* firms in other advanced economies that are capable of reproducing, replicating or reverse engineering their technology. Since few firms in low-income countries pose a similar competitive threat, establishing patent rights in these countries is a much lower priority. On the other hand, for finished products that are ready to come to market, similar firms in advanced economies apply for TMs more generally across all income groups with particular emphasis on low-income groups. This may reflect concerns about counterfeit goods in developing countries or a preference for TMs as cheaper and easier to apply for and enforce than patents. While enforcement of both patent and TM rights in developing countries may be less certain than in developed countries, the persistent prevalence of trademarking in low-income countries suggests that firms see value in TMs in these contexts. Obviously, there is more to be done in this vein – work that is only possible using industry-level concordances for TMs and patents.

FIGURE 6
Foreign Patent Intensity (per $ billion in Exports) for Various Goods and Services by Income Group, 2004–11



Note:
'Food' consists of NICE classes 29, 30 and 31. 'High-tech' consists of NICE classes 9 and 42. 'Textiles' consists of NICE classes 23, 24 and 25.

c Regression Analysis of Foreign Trademark Intensity

To go beyond the descriptive analysis in the previous subsection, we next analyse more precisely how different country and country–TM class factors influence the decision to TM in a jurisdiction. To do this, we utilise detailed country–industry-level data sets compiled by the OECD Structural Analysis (STAN) database and measure the number of TM flows to a jurisdiction based on trade flows, FDI inflows and outflows, income and other variables. Specifically, we look at the determinants of foreign TM activity based on a number of country and industry-specific variables and compare these with the determinants of foreign patenting activity to highlight some important differences in the use of different types of intellectual property abroad. Figures 5 and 6 highlight some interesting patterns and differences in foreign trademarking and patenting behaviour for three types of sectors and the intention of this analysis is to shed additional light on more sectors while gauging the sensitivity of the types of IP flow to exports and other variables. Here we follow a similar methodology to Yang and Kuo (2008) who explored trade-related influences of international patenting.

Since TMs are traditionally used for 'branding' and are related to the sale of final goods, we hypothesise that bilateral foreign TM flows will be closely related to trade flows. We propose a simple version of a bilateral TM function for country i to country j in industry k as:

$$TM_{ijkt} = f_{ijk}\left(TRADE_{ijkt}, X_{ijkt}\right), \tag{3}$$

where TM_{ijkt} is the number of TM registrations and f_{ijk} represents the function. Let $TRADE_{ijkt}$ denote all of the trade-related influences such as exports, production in k taking place in each country and trade costs, while X_{ijkt} denotes country and industry-specific characteristics. To keep the model simple, we assume that each of the components is an exponential function with a linear combination of each of their arguments so that we may log-linearise it for the estimation. We specify $TRADE_{ijkt}$ to be a function of several related variables:

$$TRADE_{ijkt} = g\left(EXP_{ijkt}, OVA_{ikt}, FVA_{jkt}, DIST_{ij}\right), \tag{4}$$

where EXP_{ijkt} total exports (value) from country i to country j in industry k at time t, OVA_{ikt} and FVA_{jkt} are the origin and foreign value-added measures for industry k, and $DIST_{ij}$ is the distance between countries i and j. In terms of the country and industry-specific arguments, we hypothesise that levels of investment, wealth and market size are the key determinants so that symbolically, we have the following:

$$X_{ijkt} = h\left(FDI_{ijt}, MARKET_{ijt}, WEALTH_{jt}, INFDI_{ikt}, OUTFDI_{jkt}, IPR_{jt}\right), \tag{5}$$

where FDI_{ijt} is bilateral FDI flows *per capita* (measured at the country level since industry-level data are not available), $MARKET_{ijt}$ is relative market size (measured as destination country's GDP divided by the origin country's GDP), $WEALTH_{ijt}$ is the relative wealth (measured as destination country's GDP *per capita* divided by the origin country's GDP *per capita*) and $INFDI_{jkt}$ and $OUTFDI_{ikt}$ are the destination country's FDI inflows and origin country's FDI outflows by industry.[11] One key addition in our analysis is the inclusion of an interaction term between the destination country's market size, $MARKET_{ijt}$, and their intellectual property rights (IPR), IPR_{jt}. The IPR variable comes from Park (2008) who assigns scores of 1–5 to

[11] Data on FDI are available at either the cross-country bilateral flows (no industry breakdown) or industry-level country inflows and outflows (no bilateralness), thereby requiring multiple measures and definitions of FDI.

country IPR strength (5 is the highest) based on a country's coverage, membership in international treaties, duration of protection, enforcement mechanisms and restrictions. The interaction term captures potential non-linearities due to IPR potentially having a larger impact in larger markets than in smaller markets (Chen and Puttitanun, 2005). Combining all of these components and taking logs produces our estimating equation as:

$$\begin{aligned} TM_{ijkt} = {} & \alpha_0 + \alpha_1 \ln\left(EXP_{ijkt}\right) + \alpha_2 \ln\left(FDI_{ijt}\right) + \alpha_3 \ln\left(MARKET_{ijt}\right) + \alpha_4 \ln\left(WEALTH_{ijt}\right) \\ & + \alpha_5 \ln\left(OVA_{ikt}\right) + \alpha_6 \ln\left(FVA_{jkt}\right) + \alpha_7 \ln\left(INFDI_{ikt}\right) + \alpha_8 \ln\left(OUTFDI_{jkt}\right) \\ & + \alpha_9 \ln\left(IPR_{jt}\right) + \alpha_{10} \ln\left(IPR_{jt}\right) \times \ln\left(MARKET_{ijt}\right) + \alpha_{10} DIST_{ij} + X_{ijt} + \delta_t + \varepsilon_{ijkt}, \end{aligned} \tag{6}$$

where X are dummy variables for additional trade costs such as contiguity and sharing the same language, and we include country and time fixed effects (δ_t, δ_i and δ_j).

To highlight specific differences in the way that intellectual property is used differently by industry and type, we estimate the same specification for patents. This specification is very similar to that of Yang and Kuo (2008), who measured outbound international patents at the country level, except that our analysis excludes R&D and schooling data and includes other variables such as industry value-added and additional measures of FDI.

Since we are measuring TM and patent flows, which are discrete variables we use a Poisson-based estimator in our analysis.[12] We first run the estimation across all 45 NICE classes and then sort the NICE classes into seven separate groups. The groups consist of 'chemicals' (NICE classes 1, 2, 3, 4 and 5), 'metals & machinery' (NICE classes 6–8, 12 and 14), 'high-tech' (NICE classes 9, 38 and 42), 'textiles' (NICE classes 22–26), 'food & beverages' (NICE classes 27–34), 'other manufacturing' (NICE classes 10, 11, 15–21, 27 and 28) and 'other services' (NICE classes 35–37, 39–41 and 43–45).

We run the analysis at the country–industry level across all OECD countries between 2004 and 2011. We then run country-level regressions across 192 nations and break them up by income pairs. We include a list of sources of data and how they are organised and converted to the NICE group in Table 2, along with summary statistics. Each bilateral country pair exchanges on average 28 TMs with each other each year and roughly 13 patents (excluding zeroes). Our estimating sample at the country–industry level consists of roughly 50,000 observations. To ensure consistency between the patent and TM data sets, we drop the observations that have missing patent or missing TM data. Tables 3 and 4 show the estimates for foreign trademarking and patenting.

A couple of key findings emerge from this estimation. We can see that international trade flows are strongly related to the transfer of intellectual property abroad, with exports being highly significant and positive for both TMs and patents. Across all NICE classes, patents appear to be consistently more sensitive, with certain types of goods being more sensitive to trade flows than others. For TMs, the estimated export elasticity is relatively low for 'other services' and 'high-tech' industries and highest for 'chemicals'. This ordering of the sensitivity of trademarking to trade may partly reflect structural differences in retail markets by industry, which can have important implications for the information content of TMs (i.e. the value of 'brand' name). For example, it seems likely that exported 'chemical' goods (which include pharmaceuticals) rely relatively heavily on TMs in retail market strategies since

[12] The Poisson is somewhat restrictive, requiring the mean and variance to be equal. However, as Santos Silva and Tenreyro (2006) demonstrate, this restriction does not bias the results in any significant way. We have run similar regressions using OLS and negative binomial estimators with qualitatively similar results.

TABLE 2
Data Sources for 2004–11

Type of Data	*Definition*	*Source*	*Original Classification*	*Mean/SD*	*No. of Distinct Observations*
Bilateral trademark (TM) flows	Number of exported TM registrations	World Intellectual Property Organization (WIPO)	NICE class	28.399 (468.07)	730,722
Bilateral patent flows	Number of exported patent grants	WIPO	4-digit International Patent Classification (IPC), converted to 4-digit Standard International Trade Classification (SITC) Rev. 2	13.008 (252.17)	345,533
Bilateral trade flows	Export values ($)	UN COMTRADE	4-digit SITC Rev. 2	16608.07 (355015.7)	3,003,971
Bilateral FDI flows	Amount of FDI exported ($)	OECD	Country level	585.301 (4618.08)	29,330
Country–industry FDI	FDI inflow and outflow positions by industry ($)	OECD	2-, 3- and 4-digit International Standardized Industrial Classification (ISIC) Rev. 3	67312.1 (226581.8)	10,841
Country GDP	Aggregate GDP ($)	World Bank	Country level	2.87E11 (1.16E12)	1.537
Country–industry VA	Aggregate Value Added by Industry ($)	OECD	2-, 3- and 4-digit ISIC Rev. 3	1.19E12 (1.23E13)	8,610
Country intellectual property rights	Country Intellectual Property Rights (1–5)	Park (2008)	No conversion	3.280 (0.089)	968
Gravity variables	Distance (km), Border, Language	CEPII	No conversion		

Notes:
All industry-level data were mapped to NICE classes using the Algorithmic Links with Probabilities (ALP) TM Concordance. For Country-Technology Patents, we used the ALP Patent Concordance to first convert IPC classification to ISIC, which were then mapped to the NICE classes.

TABLE 3
Poisson Regression of NICE-level Bilateral Trademark Flows for OECD Countries, 2004–11.
Dependent Variable is Bilateral Trademark Flows from Country i to Country j

	(1) All NICE	*(2) Chemicals*	*(3) Metals and Machinery*	*(4) High-Tech*	*(5) Textiles*	*(6) Food and Beverage*	*(7) Other Services*	*(8) Other Manufact.*
$\ln(EXP_{ijkt})$	0.289***	0.563***	0.434***	0.258***	0.432***	0.347***	0.240***	0.380***
	(0.0193)	(0.0307)	(0.0380)	(0.0235)	(0.0623)	(0.0309)	(0.0276)	(0.0275)
$\ln(FDI_{ijt})$	0.135***	0.0772***	0.0705**	0.144***	0.0441	0.0874***	0.156***	0.103***
	(0.0212)	(0.0218)	(0.0242)	(0.0240)	(0.0296)	(0.0245)	(0.0242)	(0.0240)
$MARKET_{ijt}$	−1.271*	−1.371*	−1.439*	−1.666**	−1.172**	−1.271*	−1.105*	−1.101
	(0.578)	(0.570)	(0.617)	(0.598)	(0.446)	(0.573)	(0.536)	(0.567)
$WEALTH_{ijt}$	0.202	0.194	0.176	0.247	0.367***	0.0525	0.177	0.239
	(0.124)	(0.125)	(0.146)	(0.139)	(0.111)	(0.124)	(0.109)	(0.133)
$\ln(FVA_{jkt})$	0.0474**	−0.0267	−0.0193	0.0295	0.0706	0.0405	0.0184	0.0136
	(0.0157)	(0.0230)	(0.0313)	(0.0245)	(0.0497)	(0.0223)	(0.0201)	(0.0225)
$\ln(OVA_{ikt})$	0.00670	−0.0362	−0.0885*	0.000878	−0.115	−0.0443	−0.00453	−0.0870*
	(0.0196)	(0.0273)	(0.0361)	(0.0331)	(0.0607)	(0.0286)	(0.0213)	(0.0359)
$\ln(IN\_FDI_{jkt})$	−0.0326	−0.143**	0.0229	0.0379	0.145**	0.0941*	−0.142***	−0.00234
	(0.0335)	(0.0549)	(0.0409)	(0.0433)	(0.0487)	(0.0457)	(0.0340)	(0.0645)
$\ln(OUT\_FDI_{ikt})$	0.134***	0.0223	0.0959*	0.276***	0.195***	−0.0447	0.138***	0.108*
	(0.0376)	(0.0452)	(0.0425)	(0.0540)	(0.0437)	(0.0480)	(0.0349)	(0.0496)
IPR_{jt}	−2.803	−3.675*	−3.298	−2.730	−4.027**	−2.738	−1.825	−3.255*
	(1.620)	(1.565)	(1.722)	(1.833)	(1.383)	(1.587)	(1.648)	(1.605)
$MARKET_{ijt} \times IPR_{jt}$	0.791*	0.814*	0.904*	1.100**	0.654*	0.713	0.726*	0.644
	(0.382)	(0.374)	(0.407)	(0.397)	(0.294)	(0.378)	(0.356)	(0.375)
log distance	0.000332	0.210*	0.0786	0.0764	0.157	−0.0540	−0.0647	0.0877
	(0.0800)	(0.0870)	(0.0954)	(0.0936)	(0.116)	(0.0782)	(0.0895)	(0.0897)
Border dummy	0.323	0.259	0.248	0.419*	0.112	0.338*	0.270	0.243
	(0.174)	(0.184)	(0.192)	(0.207)	(0.140)	(0.164)	(0.197)	(0.178)
Language	0.326	−0.0767	0.0811	0.427*	0.285	0.278	0.606***	0.273
Dummy	(0.177)	(0.173)	(0.205)	(0.197)	(0.146)	(0.181)	(0.177)	(0.184)
Constant	−2.038	−2.260	−1.037	−3.855	−3.768	−0.744	0.100	−0.906
	(2.524)	(2.490)	(2.832)	(2.875)	(2.242)	(2.560)	(2.489)	(2.580)
Observations	50,163	6,115	4,909	4,043	4,338	6,872	9,551	14,335

Notes:
(i) All estimations use year fixed effects. Estimation (1) also contains NICE class fixed effects.
(iI) Robust standard errors in parentheses clustered by bilateral pairs, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

TABLE 4
Poisson Regression of NICE-level Bilateral Patent Flows for OECD Countries, 2004–11.
Dependent Variable is Bilateral Patent Flows from Country i to Country j

	(1) *All NICE*	(2) *Chemicals*	(3) *Metals and Machinery*	(4) *High-tech*	(5) *Textiles*	(6) *Food and Beverage*	(7) *Other Services*	(8) *Other Manufact.*
$\ln(EXP_{ijkt})$	0.735***	0.637***	0.554***	0.861***	0.495***	0.432***	0.462***	0.567***
	(0.0447)	(0.0658)	(0.0757)	(0.0436)	(0.0813)	(0.0473)	(0.0678)	(0.102)
$\ln(FDI_{ijt})$	0.212***	0.184**	0.291***	0.241***	0.308***	0.345***	0.284***	0.341***
	(0.0517)	(0.0573)	(0.0797)	(0.0703)	(0.0548)	(0.0551)	(0.0687)	(0.0992)
$MARKET_{ijt}$	−3.843***	−3.861***	−3.501***	−5.401***	−2.108*	−2.823***	−3.729***	−3.741***
	(0.800)	(0.845)	(1.042)	(1.211)	(1.050)	(0.790)	(0.899)	(1.027)
$WEALTH_{ijt}$	−0.186	0.0910	−0.394*	−0.380	−0.449*	−0.649**	−0.336*	−0.402*
	(0.189)	(0.196)	(0.160)	(0.209)	(0.206)	(0.213)	(0.155)	(0.187)
$\ln(FVA_{jkt})$	0.0811*	0.165***	0.165***	0.203***	0.195***	0.0251	0.122**	0.155***
	(0.0360)	(0.0431)	(0.0320)	(0.0316)	(0.0436)	(0.0376)	(0.0442)	(0.0277)
$\ln(OVA_{ikt})$	0.177**	0.199***	0.196***	0.200***	0.172**	0.207**	0.216***	0.231**
	(0.0570)	(0.0384)	(0.0438)	(0.0488)	(0.0643)	(0.0720)	(0.0527)	(0.0743)
$\ln(IN\_FDI_{jkt})$	−0.117	0.106	0.0135	−0.113	0.138	−0.0427	−0.158***	−0.210
	(0.0879)	(0.107)	(0.0640)	(0.108)	(0.0770)	(0.102)	(0.0360)	(0.142)
$\ln(OUT\_$	0.0928	0.0410	−0.120*	0.186*	0.00585	−0.246**	0.0688	−0.0945
$FDI_{ikt})$	(0.0578)	(0.0676)	(0.0526)	(0.0784)	(0.0694)	(0.0792)	(0.0461)	(0.0843)
IPR_{jt}	7.810*	3.049	15.80***	9.027	9.971*	11.39*	10.63**	14.35**
	(3.980)	(3.954)	(4.316)	(4.809)	(4.136)	(4.730)	(4.028)	(5.094)
$MARKET_{ijt}$	2.555***	2.452***	2.209**	3.664***	1.274	1.769***	2.467***	2.425***
$\times\ IPR_{jt}$	(0.520)	(0.533)	(0.682)	(0.782)	(0.678)	(0.500)	(0.579)	(0.671)
log distance	0.517*	0.710***	0.450*	0.258	0.507**	0.624**	0.399*	0.414*
	(0.237)	(0.146)	(0.180)	(0.170)	(0.193)	(0.234)	(0.192)	(0.187)
Border	−0.133	0.730	0.105	−0.716	0.433	0.382	0.122	−0.0719
dummy	(0.587)	(0.414)	(0.487)	(0.457)	(0.457)	(0.597)	(0.478)	(0.534)
Language	−0.105	−0.340	−0.165	0.0158	−0.250	−0.111	−0.244	−0.116
Dummy	(0.178)	(0.274)	(0.158)	(0.171)	(0.249)	(0.263)	(0.207)	(0.207)
Constant	−37.32***	−33.54***	−47.10***	−44.74***	−40.67***	−35.70***	−38.34***	−44.00***
	(5.457)	(5.950)	(6.054)	(6.426)	(6.243)	(6.787)	(5.451)	(7.022)
Observations	50,163	6,115	4,909	4,043	4,338	6,872	9,551	14,335

Notes:
(i) All estimations use year fixed effects.
(ii) Robust standard errors in parentheses clustered by bilateral pairs, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

consumers would tend to rely more on the reputational information of the brand in order to infer the quality of the product as opposed to generic products. Following this logic, industries with greater information asymmetries between the consumer and seller are likely to use TMs more heavily in order to reduce search costs, as described in Economides (1987) and Landes and Posner (1987).

Turning to the patenting results, Table 4 indicates that international patenting in 'High-tech' goods is the most sensitive to exports relative to other goods. 'Chemicals' also has a much higher export elasticity than other goods. Both these industries are known to rely heavily on patents (Taylor and Silberston, 1973; Mansfield, 1986; Granstrand, 1999). On the other hand, industries such as 'food & beverages', 'textiles' and 'other services' rely much less on patents (Granstrand, 1999), which is also evident in these patenting results.

Since patenting and trademarking are both strategies to protect intellectual property, it is potentially insightful to visualise their respective sector-level export elasticities jointly. Figure 7 plots each sector's export elasticity estimates for international trademarking and patenting along with 95 per cent confidence intervals. Pooling all sectors together produces substantial and statistically significant heterogeneity by sector. Patents tend to be more sensitive to exports than TMs with the largest differences between the two occurring in 'high-tech' goods, where exports have an elasticity nearly four times higher for patents than for TMs. In addition to this 'innovation sensitivity' story, Figure 7 captures an important 'branding sensitivity' dimension as well, which among other things reflects the retail information asymmetries described above. In 'foods and beverages' and 'textiles', the impact of trade on flows of brands is just as important as new innovations (as measured in patents).

Looking at the other explanatory variables, we find that bilateral FDI flows can also partially explain TM patterns and that IPR has a negative (albeit mostly insignificant) effect as a standalone variable. Given that the sample consists of mostly wealthy countries with mostly high levels of IPR, this result should not be interpreted too strongly. In the interaction term, we find that IPR combined with relative market size has a significant positive effect, supporting our hypothesis that IPR rights are more important in larger countries.

Turning now to the country-level analysis, we assess how income differences affects intellectual property flows across similar explanatory variables. Our observations include countries outside of the OECD. Table 5 highlights the results. Across all countries, we find that exports have an elasticity of approximately 0.43; with relative wealth negatively effecting bilateral TM flows.

FIGURE 7
Export Elasticity of Patents versus Trademarks

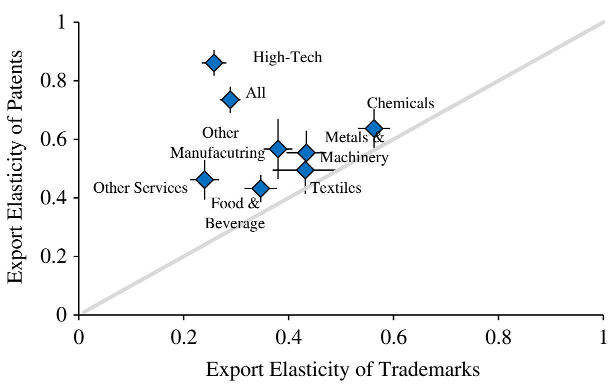


TABLE 5
Poisson Regression of Country-level Bilateral Trademark (TM) Flows by Income Level, 2004–11.
Dependent Variable is Bilateral TM Flows from Country *i* to Country *j*

	(1) All Countries	*(2) Low-to-Low*	*(3) Low-to-Middle*	*(4) Low-to-High*	*(5) Middle-to-Low*	*(6) Middle-to-Middle*	*(7) Middle-to-High*	*(8) High-to-Low*	*(9) High-to-Middle*	*(10) High-to-High*
$\ln(EXP_{ijkt})$	0.430***	0.0991	0.245***	0.140***	0.212***	0.341***	0.397***	0.207***	0.338***	0.477***
	(0.0272)	(0.0539)	(0.0484)	(0.0237)	(0.0415)	(0.0273)	(0.0208)	(0.0362)	(0.0287)	(0.0189)
$MARKET_{ijt}$	0.182	−0.538	0.484	−0.00250	−1.185***	−0.0752	−0.998***	−0.740***	0.453*	−0.798***
	(0.128)	(0.512)	(0.307)	(0.649)	(0.179)	(0.222)	(0.199)	(0.141)	(0.229)	(0.132)
$WEALTH_{ijt}$	−0.200***	−0.207	0.0503	−0.621***	0.283**	0.101	−0.0812	−0.102	−0.0237	0.0788
	(0.0383)	(0.151)	(0.119)	(0.0596)	(0.0996)	(0.0967)	(0.0424)	(0.0767)	(0.0727)	(0.0444)
$\ln(IN\_FDI_{jt})$	0.313***	1.283***	0.471***	0.467***	0.772***	0.423***	0.564***	0.882***	0.437***	0.135***
	(0.0423)	(0.174)	(0.0969)	(0.0672)	(0.106)	(0.0748)	(0.0312)	(0.0846)	(0.0644)	(0.0237)
$\ln(OUT\_FDI_{it})$	0.213***	0.228**	0.323***	0.339***	0.0749	0.193***	−0.00927	0.357***	0.475***	0.216***
	(0.0330)	(0.0853)	(0.0581)	(0.0372)	(0.0465)	(0.0355)	(0.0231)	(0.0491)	(0.0366)	(0.0210)
IPR_{jt}	0.860*	4.040***	0.713	4.081***	2.977***	1.634***	3.439***	3.334***	2.060***	0.215
	(0.369)	(0.835)	(0.597)	(0.688)	(0.702)	(0.450)	(0.441)	(0.437)	(0.421)	(0.320)
$MARKET_{ijt}$	−0.299***	−0.0696	−0.448*	−0.0594	0.431**	−0.122	0.295*	0.235*	−0.450**	0.413***
$\times\ IPR_{jt}$	(0.0906)	(0.447)	(0.223)	(0.415)	(0.146)	(0.167)	(0.130)	(0.119)	(0.170)	(0.0896)
log distance	−0.207***	−1.384***	−1.338***	−1.093***	−1.485***	−0.755***	−0.716***	−1.122***	−0.650***	−0.0244
	(0.0445)	(0.167)	(0.0964)	(0.0862)	(0.0754)	(0.0680)	(0.0371)	(0.0747)	(0.0381)	(0.0247)
Border	0.110	−0.0443	−0.0912	−0.0611	−0.364*	0.0942	−0.181	0.139	−0.401*	0.385***
dummy	(0.137)	(0.420)	(0.194)	(0.137)	(0.151)	(0.145)	(0.0947)	(0.214)	(0.187)	(0.0686)
Language	0.0282	−0.763*	−0.00655	0.324*	−2.254***	0.180	−0.465***	−0.608***	−0.267*	0.196***
Dummy	(0.111)	(0.321)	(0.210)	(0.130)	(0.294)	(0.101)	(0.0800)	(0.170)	(0.126)	(0.0551)
Constant	−10.22***	−9.109***	−0.449	−4.332***	−1.481	−6.037***	−10.05***	−9.155***	−10.32***	−9.598***
	(0.828)	(1.229)	(1.267)	(1.204)	(1.379)	(0.859)	(0.730)	(1.196)	(0.730)	(0.583)
Observations	59,453	6,374	4,869	7,062	5,639	4,494	6,521	8,201	6,913	9,380

Notes:
(i) All estimations use year fixed effects.
(ii) Robust standard errors in parentheses clustered by bilateral pairs, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

We find that the destination country's inward FDI and the origin country's outward FDI are positively associated with TMs, along with IP protection. Interestingly, the interactive term of relative market size and IPR have a negative effect at this country level, suggesting that IP protection in larger markets is less important. Looking at the breakdown by income levels, we find that the negative effect is being mostly driven by middle-income destination markets, suggesting that TM applicants care more about the relative market size versus IPR in these countries. This may be due to the expectation that IPR will evolve as these countries transition from low to high income. Of the gravity variables, only distance is significant and negative.

Across income pairs, we find substantial differences for several of the coefficients. Trade elasticities for TMs are highest in high and middle-income destination countries, with the elasticities being 1.5–2× higher than when the destination country is a low-income country. TMs flowing to low-income destinations appear to be mostly driven by investment levels (Inward FDI) entering the low-income destination. Distance also seems to be nearly twice as strong a determinant of TM flows from firms in middle and low-income countries as from firms in high-income countries. Indeed, distance plays no role in TM flows between high-income countries, but border effects and language are both positive and significant. We estimated the same specification for patents (see Appendix A Table A3). We find similar signs across the variables, with the exception of border effects, which are negative for patents. Among high-income-to-high-income transfers, we see that distance actually has a significantly positive effect for patents, perhaps indicating that 'geographic coverage' is more important for patents than for TMs.

In sum, we see some pronounced differences across both income and industry dimensions in the how patents and TMs are used internationally and how responsive they are to trade flows. These patterns likely reflect structural differences in production and demand between industries and institutional differences across countries in different income groups. Mapping more precisely these patterns and differences in specific industrial clusters and regional groups is a promising area for future research.

d. Robustness Check

One key aspect of the estimation to consider is the fact that it is likely that both trade and FDI flows are heavily influenced by gravity measures, such as market size, distance and more. Thus, it may be the case that equation (3) suffers from endogeneity issues where trade and FDI are correlated with the error term. As an additional check, we run an instrumental variable estimation where *TRADE* and *FDI* are instrumented using the origin and destination country's GDP, value-added, distance, border and language dummies.[13] These results can be found in Table 6, which include the first and second stages.

The first stage lends support for endogeneity of trade and FDI flows, as both are heavily influenced by gravity variables. Once we account for the endogeneity, we find that this does not significantly alter the estimated trade elasticities for either patents or TMs. We run the same specification across the sectors and report the second-stage results in Table 7. Here we find mostly similar results, with the exception of 'textiles' whose coefficient values differ significantly from every other sector. In the 'textiles' case, trade is negatively associated with

[13] Specifically, we regress: $\ln(EXP_{ijkt}) = \beta_0 + \beta_1 \ln(OGDP_{it}) + \beta_2 \ln(FGDP_{jt}) + \beta_3 \ln(OVA_{ikt}) + \beta_4 \ln(FVA_{jkt}) + \beta_5 \ln(DIST_{ij}) + BORDER_{ij} + LANG_{ij}$ and $\ln(FDI_{ijt}) = \beta_0 + \beta_1 \ln(OGDP_{it}) + \beta_2 \ln(FGDP_{jt}) + \beta_3 \ln(DIST_{ij}) + BORDER_{ij} + LANG_{ij}$.

TABLE 6
IV Regression of Bilateral Trademark (TM) and Patent Flows for OECD Countries, 2004–11. Dependent Variable is Bilateral TM Flows from Country i to Country j

	First Stage		*Second Stage*	
	(1) $ln(EXPORTS_{ijkt})$	*(2)* $ln(FDI_{ijt})$	*(3)* *TMs – All NICE*	*(4)* *Patents – All NICE*
$\ln(OGDP_{it})$	0.840***	1.047***		
	(0.0325)	(0.0619)		
$\ln(FGDP_{jt})$	0.789***	0.884***		
	(0.0259)	(0.0588)		
$\ln(\widehat{EXP}_{ijkt})$			0.206***	0.937***
			(0.0142)	(0.0278)
$\ln(FDI_{ijt})$			0.244***	0.316**
			(0.0483)	(0.0963)
$MARKET_{ijt}$			−0.942	−3.579***
			(0.566)	(0.948)
$WEALTH_{ijt}$			0.149	−0.467*
			(0.103)	(0.232)
$\ln(FVA_{jkt})$	0.0262	−0.0960**	0.0573**	0.0686
	(0.0154)	(0.0352)	(0.0196)	(0.0352)
$\ln(OVA_{ikt})$	0.0503**	−0.134***	0.0270	0.256***
	(0.0158)	(0.0378)	(0.0188)	(0.0568)
$\ln(IN\_FDI_{jkt})$			0.0288	−0.163*
			(0.0354)	(0.0633)
$\ln(OUT\_FDI_{ikt})$			0.0705	0.0740
			(0.0386)	(0.0537)
IPR_{jt}			−2.138	2.396
			(1.875)	(3.521)
$MARKET_{ijt} \times IPR_{jt}$			0.530	2.540***
			(0.378)	(0.624)
log distance	−1.006***	−0.997***	0.0598	1.141***
	(0.0402)	(0.0802)	(0.106)	(0.253)
Border dummy	0.649***	−0.458*	0.388	−0.172
	(0.138)	(0.225)	(0.245)	(0.655)
Language dummy	0.668***	1.002***	0.0901	−0.693*
	(0.121)	(0.203)	(0.195)	(0.280)
Constant	−28.36***	−33.66***	−3.361	−39.23***
	(1.075)	(2.055)	(2.582)	(5.124)
F-statistic	2011.7	19.09		
R^2	0.888	0.495		
Observations	72,428	72,428	72,428	72,428

Notes:
(i) All estimations use year fixed effects. First stage is estimated using a panel fixed-effects model.
(ii) Robust standard errors in parentheses clustered by bilateral pairs, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

TM flows, while bilateral FDI flows are the most significant determinant. This may be due to the peculiar industry structure of textiles which sees most trade flows going from low income to high income and most of the branding occurring in high-income countries. We also find that aggregate inward FDI for the destination country and outward FDI for the origin country are also positive and significant, along with value-added and relative wealth.

TABLE 7
IV Regression of NICE-level Bilateral Trademark (TM) Flows for OECD Countries, 2004–11.
Dependent Variable is Bilateral TM Flows from Country i to Country j

	(1) All NICE	(2) Chemicals	(3) Metals and Machinery	(4) High-tech	(5) Textiles	(6) Food and Beverage	(7) Other Services	(8) Other Manufact.
$\ln(EXP_{ijkt})$	0.206***	0.959***	0.128***	0.207***	−2.921***	0.432***	0.251***	0.215***
	(0.0142)	(0.0438)	(0.0350)	(0.0204)	(0.207)	(0.0177)	(0.0127)	(0.0175)
$\ln(FDI_{ijt})$	0.244***	−0.416***	0.317***	0.213***	2.484***	−0.0132	0.198***	0.271***
	(0.0483)	(0.0650)	(0.0595)	(0.0531)	(0.173)	(0.0475)	(0.0517)	(0.0559)
$MARKET_{ijt}$	−0.942	−0.727	−1.010	−1.367*	−0.481	−0.808	−0.734	−0.802
	(0.566)	(0.578)	(0.682)	(0.583)	(0.527)	(0.509)	(0.537)	(0.583)
$WEALTH_{ijt}$	0.149	0.0653	0.0969	0.180	0.528***	0.0910	0.142	0.196
	(0.103)	(0.105)	(0.125)	(0.118)	(0.117)	(0.0946)	(0.0948)	(0.114)
$\ln(FVA_{jkt})$	0.0573**	−0.0813*	0.0355	0.0210	0.314***	0.0517*	0.0269	0.0384
	(0.0196)	(0.0359)	(0.0405)	(0.0312)	(0.0487)	(0.0252)	(0.0250)	(0.0268)
$\ln(OVA_{ikt})$	0.0270	−0.106**	0.00534	0.0274	0.440***	−0.0180	−0.0110	−0.0429
	(0.0188)	(0.0343)	(0.0376)	(0.0319)	(0.0530)	(0.0270)	(0.0206)	(0.0324)
$\ln(IN\_FDI_{jkt})$	0.0288	0.0160	0.0832	0.0685	0.538***	0.199***	−0.102***	0.111
	(0.0354)	(0.0673)	(0.0496)	(0.0416)	(0.0637)	(0.0513)	(0.0269)	(0.0704)
$\ln(OUT\_$	0.0705	0.0525	0.0208	0.232***	0.373***	−0.107*	0.0847**	0.0353
$FDI_{ikt})$	(0.0386)	(0.0585)	(0.0551)	(0.0413)	(0.0559)	(0.0537)	(0.0283)	(0.0594)
IPR_{jt}	−2.138	−2.472	−2.127	−1.565	−6.642***	−2.026	−0.831	−3.134
	(1.875)	(1.747)	(1.983)	(2.151)	(1.736)	(1.610)	(2.097)	(1.864)
$MARKET_{ijt}$	0.530	0.357	0.580	0.869*	0.353	0.350	0.428	0.399
$\times IPR_{jt}$	(0.378)	(0.381)	(0.453)	(0.390)	(0.347)	(0.339)	(0.361)	(0.387)
log distance	0.0598	0.303**	0.106	0.160	−0.759***	−0.0339	0.0107	0.137
	(0.106)	(0.113)	(0.115)	(0.123)	(0.110)	(0.0968)	(0.130)	(0.110)
Border	0.388	−0.0498	0.580**	0.455	3.506***	0.226	0.185	0.441
dummy	(0.245)	(0.236)	(0.225)	(0.280)	(0.306)	(0.240)	(0.299)	(0.239)
Language	0.0901	−0.155	−0.166	0.235	−0.618**	0.0565	0.364	−0.0134
Dummy	(0.195)	(0.202)	(0.218)	(0.218)	(0.209)	(0.185)	(0.194)	(0.202)
Constant	−3.361	−7.907**	−2.079	−6.015*	26.51***	−4.063	−2.311	−1.472
	(2.582)	(2.579)	(2.882)	(2.964)	(2.792)	(2.269)	(2.751)	(2.619)
Observations	72,428	8,893	7,073	5,952	6,097	9,837	13,875	20,701

Notes:
(i) All estimations use year fixed effects. Estimation (1) also contains NICE class fixed effects.
(ii) Robust standard errors in parentheses clustered by bilateral pairs, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

5. CONCLUSIONS AND FUTURE WORK

To summarise, we have explored differences in behaviour of firms taking out intellectual property abroad using count estimation and an IV approach. This analysis points to several subtle differences in the behaviour of industries that patent and TM internationally. We have shown that across all industries, TMs appear to be less sensitive to trade flows relative to patents, with some substantial differences across sectors. We have developed a very basic theoretical justification for the underlying specification as the key point in our analysis is to illustrate the potential use of a NICE class–industry concordance to enhance our understanding of technology transfer and more. More detailed analysis needs to be carried out to investigate the causes for these differences, which we leave to future research.

Although TMs are fundamentally important to business strategy and market efficiency in many sectors, economists and policy analysts are severely constrained when it comes to empirical options. TM data and economic activity data are organised differently and could not be analysed jointly at a level of resolution that matches the marked heterogeneity in how TMs are used in different industries. This paper describes our attempt to remedy this incompatibility by building a bridge between TM and economic data – a bridge that can support analyses that are far more disaggregated than previously possible.

To build this linkage, we develop an algorithmic approach we call ALP matching, which we originally designed for applications to patent data. ALP matching generates TM-specific links to trade and industry classifications and processes these raw matches into aggregate concordances. Specifically, NICE class-level concordances can map TM data into trade or industry categories, or, alternatively, trade or industry data into NICE classes. As a key benefit to this approach, these class-level ALP concordances implicitly reflect differences in TM usage across economic sectors – and therefore link TMs to economic activity according to predominant TM-use patterns. We demonstrate the use of this approach via numerous sample analyses of countries in which we depict differences in TM-use intensity and compare foreign TM and patent use. There is much more that could be done with linked TM-economic activity data depending on one's research objective. Since TM flows can now be organised by different industry classification systems, we can merge other data sets organised by these classification systems, such as trade elasticities of substitution from Broda and Weinstein (2006) and the Rauch classification of goods (Rauch, 1999). It is also possible to account for the number of varieties within product sectors using TM data, which can now be used to test outcomes predicted in various trade and growth models, such as welfare measures from an expanding or shrinking variety set. In this paper, we highlighted several ways in which industries utilise intellectual property differently, with more work to follow.

Several dimensions of future work related to this paper are worth noting. First, while we believe the ALP approach to constructing concordances represents a valuable contribution to research in this area, there is surely more that could be done to refine these algorithmic methods. The approaches we describe above generate ALP concordances at the NICE class level. The aggregation involved in constructing these concordances has the advantage of sweeping away the noise that is inevitable in the probabilistic match of an individual TM to economic categories. For descriptive exercises that seek to compare TM and economic landscapes, class-level concordances provide a sufficient degree of disaggregation without unnecessary detail. For more rigorous statistical modelling of TMs, greater resolution may be more useful. One could, for example, use ALP matching to directly classify a specific TM according to economic classifications. Alternatively, a GS indicator-level concordance may provide a convenient middle ground between class-level concordances and TM-specific linkages. Given the difficulties described above inherent in working with the NICE classification system, either of these approaches presents some considerable challenges.

Second, until yet more sophisticated ALP approaches are devised, we believe significant future work could be based on the class-level ALP concordances as they now exist. There are several promising ways to use these concordances to push beyond descriptive analysis. Since different sectors and industries use TMs quite differently, the impact of this form of IP on our contemporary economy is distinctly heterogeneous. Consequently, just about any analysis of TMs in the modern economy will be empirically richer and more insightful once TM data and economic data can be jointly analysed at disaggregated levels.

REFERENCES

Allegrazza, S. and A. Guard-Rauchs (1999), 'The Determinants of Trademark Deposits: An Econometric Investigation (A Case Study of the Benelux)', *Economie Applique*, **52**, 2, 51–68.

Broda, C. and D. Weinstein (2006), 'Globalization and the Gains from Variety', *Quarterly Journal of Economics*, **121**, 2, 541–85.

Chen, Y. and T. Puttitanun (2005), 'Intellectual Property Rights and Innovation in Developing Countries', *Journal of Development Economics*, **78**, 2, 474–93.

Economides, N. (1987), 'The Economics of Trademarks', *Trademark Register*, **78**, 523–39.

Fink, C., B. Javorcik and M. Spatareanu (2003), 'Income-Related Biases in International Trade: What do Trademark Registration Data Tell Us?', *Review of World Economics*, **141**, 1, 79–103.

Granstrand, O. (1999), 'Internationalization of Corporate R&D: A Study of Japanese and Swedish Corporations', *Research Policy*, **28**, 23, 275–302.

Greenhalgh, C. and M. Rogers (2012), 'Trade Marks and Performance in UK Firms: Evidence of Schumpeterian Competition Through Innovation', *Australian Economic Review*, **45**, 1, 50–76.

Greenhalgh, C., M. Longland and D. Bosworth (2001), 'Protecting Intellectual Property: British, European and American Patents and Trademarks of Selected UK Companies, 1986–1995', OIPRC Working Paper (Oxford: OIPRC).

Jensen, P. and E. Webster (2004), 'Recent Patterns of Trademarking Activity in Australia', *Australian Journal of Intellectual Property*, **15**, 2, 112–26.

Landes, W. and R. Posner (1987), 'Trademark Law: An Economic Perspective', *Journal of Law and Economics*, **30**, 2, 265–309.

Loundes, J. and M. Rogers (2003), 'The Rise of Trade Marking in Australia in the 1990s', Melbourne Institute Working Paper 8 (Melbourne: Melbourne Institute).

Lybbert, T. and N. Zolas (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**, 2, 530–42.

Mainwaring, L., N. Moore and P. Murphy (2004), 'Trademark Holdings of Production Firms in Britain and Ireland', Working Paper (Swansea: University of Wales, Swansea).

Malmberg, C. (2005), 'Trademarks Statistics as Innovation Indicator? – A Micro Stud', Centre for Innovation, Research and Competence in the Learning Economy Working Paper (Lund: CIRCLE).

Mangani, A. (2007), 'Measuring Variety and Quality of Products with Trademarks', *International Economic Journal*, **21**, 4, 613–31.

Mansfield, E. (1986), 'Patents and Innovation: An Empirical Study', *Management Science*, **32**, 2, 173–81.

Mendonca, S., T. Pereira and M. Godinho (2004), 'Trademarks as an Indicator of Innovation and Industrial Change', *Research Policy*, **33**, 9, 1385–404.

Millot, V. (2009), 'Trademarks as an Indicator of Product and Marketing Innovations', OECD Science, Technology and Industry Working Paper (Paris: OECD).

Park, W. G. (2008), 'International Patent Protection: 1960–2005', *Research Policy*, **37**, 4, 761–6.

Ramello, G. (2006), 'What's in a Sign? Trademark Law and Economic Theory', *Journal of Economic Surveys*, **20**, 4, 547–65.

Rauch, J. (1999), 'Networks versus Markets in International Trade', *Journal of International Economics*, **48**, 1, 7–35.
Rujas, J. (1999), 'Trademarks: Complementary to Patents', *World Patent Information*, **21**, 1, 35–39.
Santos Silva, J. M. C. and S. Tenreyro (2006), 'The Log of Gravity', *The Review of Economics and Statistics*, **88**, 4, 641–58.
Schautschick, P. and C. Greenhalgh (2013), 'Empirical Studies of Trade Marks – The Existing Economic Literature', Working Paper (Oxford: Oxford University).
Scherer, F. M. (1983), 'The Propensity of Patent', *International Journal of Industrial Organization*, **1**, 1, 107–28.
Schmoch, U. (2003), 'Services Marks as Novel Innovation Indicator', *Research Evaluation*, **12**, 2, 149–56.
Taylor, C. T. and A. Silberston (1973), *The Economic Impact of the Patent System: A Study of the British Experience* (Cambridge: Cambridge University Press).
WIPO (2012), 'World intellectual property indicators', Available at: http://www.wipo.int/edocs/pubdocs/en/intproperty/941/wipo_pub_941_2012.pdf (accessed 23 January 2016).
Yang, C.-H. and N.-F. Kuo (2008), 'Trade-related Influences, Foreign Intellectual Property Rights and Outbound International Patenting', *Research Policy*, **37**, 3, 446–59.
Yorukoglu, M. (2000), 'Product vs. Process Innovations and Economic Fluctuations', *Carnegie-Rochester Conference Series on Public Policy*, **52**, 1, 137–63.

APPENDIX A

a. Additional Details Regarding the ALP Matching Methodology

(i) Comparison of USPTO-Based and ROMARIN-Based Class-level ALP Concordances

The process we devised for constructing the Algorithmic Links with Probabilities (ALP) concordances described above can take any TM data as raw input. If TM use patterns vary systematically from one jurisdiction to another, then differences in TM usage could translate into differences in concordances based on these data. Specifically, if TMs in a particular jurisdiction are infrequently registered in a few NICE classes, then ALP matching may not be very robust for these classes. We avoid these potential problems using two TM databases – USPTO and ROMARIN – with substantial TM activity across all NICE classes.

As mentioned above, we expect *a priori* ALP concordances based on USPTO data to be less noisy than those based on ROMARIN data since the latter include widespread defensive registrations (which, in principle, add noise). In addition, much of the noise will be dissipated across the vast number of trademarks analysed (approximately 3.3M TMs are analysed in the USPTO data versus approximately 375K in the ROMARIN data).

We constructed class-level concordances for ISIC, HS, SITC and NAICS using both USPTO and ROMARIN data and compare these concordances to better understand potential differences. In this subsection, we directly compare the resulting ALP concordances for all industry classifications. Table A1 provides the correlations in both directions (industry-to-NICE and NICE-to-industry) of a USPTO-based crosswalk versus a ROMARIN-based crosswalk. We focus on one of each of the industry classifications and break up the comparison where we compare all weights (includes all NICE-to-industry combinations (and vice versa)), USPTO non-zeroes (all NICE-to-industry combinations (and vice versa) where weight for USPTO-based concordance is greater than zero), ROMARIN non-zeroes (all NICE-to-industry combinations (and vice versa) where weight for ROMARIN-based concordance is greater than zero), and both non-zeroes (all NICE-to-industry (and vice versa) combinations where weights for both ROMARIN-based and USPTO-based concordance is greater than zero).

TABLE A1
Correlation of Crosswalk Weights between USPTO-based ALP Concordance versus ROMARIN-based ALP Concordance

Industry Group	*Both Include Zeroes*	*USPTO Non-zeroes*	*ROMARIN Non-zeroes*	*Both Non-zeroes*
Industry-to-NICE Crosswalk				
All	0.866	0.814	0.801	0.776
Observations	11,070	790	782	538
International Standardized Industrial Classification (ISIC) Rev. 3	0.785	0.702	0.689	0.648
Observations	2,700	221	209	144
Standard International Trade Classification (SITC) Rev. 3	0.890	0.845	0.845	0.819
Observations	2,970	198	214	139
North American Industry Classification System (NAICS) 2002	0.905	0.877	0.876	0.873
Observations	1,080	102	88	66
Harmonized System (HS) 2002	0.887	0.839	0.816	0.793
Observations	4,320	269	271	189
NICE-to-Industry crosswalk				
All	0.850	0.790	0.795	0.776
Observations	11,070	703	732	501
ISIC Rev. 3	0.771	0.700	0.688	0.674
Observations	2,700	202	180	128
SITC Rev. 3	0.870	0.811	0.829	0.797
Observations	2,970	165	199	123
NAICS 2002	0.846	0.785	0.788	0.779
Observations	1,080	150	148	109
HS 2002	0.915	0.876	0.883	0.865
Observations	4,320	186	206	142

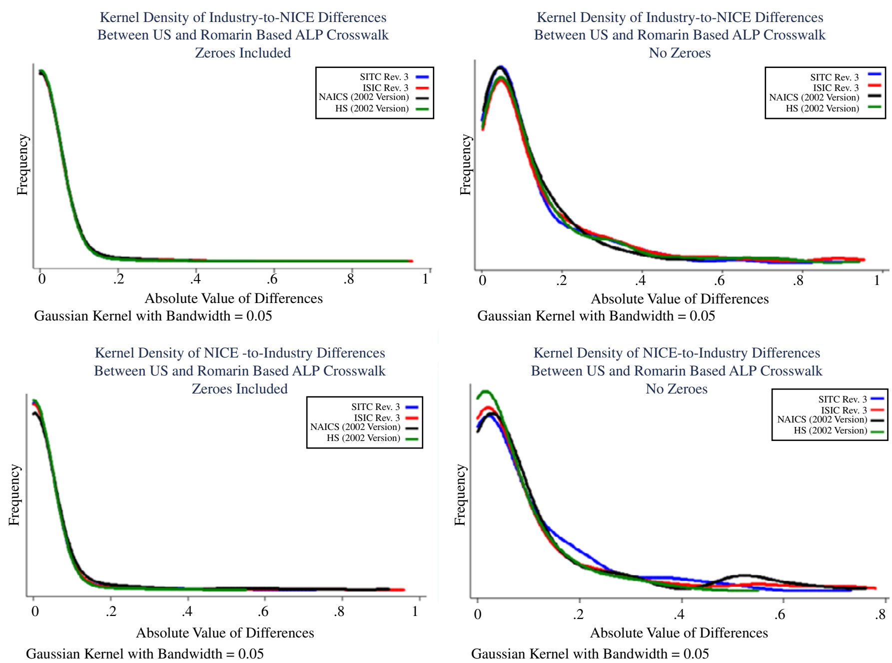
Overall, the USPTO-based and ROMARIN-based crosswalks are very similar. When we include the zero weights, the overall correlation across all industries is 0.866 for the industry-to-NICE crosswalk and 0.85 for the NICE-to-industry crosswalk, with a range of 0.771–0.915 for the worst and best matched industry classifications. When we focus on the non-zero weights in the USPTO-based classification and the ROMARIN-based classification, we see similarly high correlations. In instances where the weights in both USPTO and ROMARIN-based classifications are non-zero, we have an overall correlation of 0.776 for both directions of the crosswalk.

When we look at the value of the differences in the weights between the USPTO and ROMARIN-based crosswalks, we also see low levels. Figure A1 plots the kernel densities of the absolute differences.

The vast majority of differences between the two concordances are less than 0.1, even when we look at the non-zero weights, highlighting the consistency of the methodology and how it can be applied to other data samples.

For our actual crosswalk, we elect to stick with the USPTO-based ALP crosswalk due to the defensive nature of the ROMARIN-based crosswalk, large sample size of the USPTO TMs (more than 10 times as many TMs are analysed) and longer time horizon.

FIGURE A1
Kernel Density Plot of Absolute Differences in Weights between USPTO and ROMARIN-based ALP Concordances



b. Predefined Country Groups

TABLE A2
List of Countries by Income Group (by ISO Country-code)

Low-income Countries	*Middle-income Countries*	*High-income Countries*
AFG, ALB, ARM, BDI, BEN,	AGO, ARG, ASM, ATG, AZE,	ABW, AND, ARE, AUS,
BFA, BGD, BLZ, BOL, BTN,	BGR, BIH, BLR, BRA, BWA,	AUT, BEL, BHR, BMU,
CAF, CIV, CMR, COD, COG,	CHL, CHN, COL, CRI, CUB,	BRB, BRN, CAN, CHE,
COM, CPV, DJI, EGY, ERI,	DMA, DOM, DZA, ECU, GAB,	CUW, CYM, CYP, CZE,
ETH, FJI, GEO, GHA, GIN,	GRD, JAM, JOR, KAZ, LBN,	DEU, DNK, ESP, EST, FIN,
GMB, GNB, GTM, GUY, HND,	LBY, LTU, LVA, MDV, MEX,	FRA, GBR, GNQ, GRC,
HTI, IDN, IND, IRQ, KEN,	MNE, MUS, MYS, NAM, PAN,	GRL, GUM, HRV, HUN,
KGZ, KHM, KIR, LBR, LKA,	PER, PLW, ROU, RUS, SRB,	IMN, IRL, ISL, ISR, ITA,
LSO, MAR, MDG, MHL, MLI,	SUR, SYC, THA, TKM, TUN,	JPN, KWT, LIE, LUX,
MMR, MNG, MOZ, MRT,	TUR, TUV, URY, ZAF, IRN,	MCO, MLT, MNP, NCL,
MWI, NER, NGA, NIC, NPL,	MKD, LCA, VCT, VEN	NLD, NOR, NZL, OMN,
PAK, PHL, PNG, PRK, PRY,		POL, PRI, PRT, PYF, QAT,
RWA, SDN, SEN, SLB, SLE,		SAU, SGP, SMR, SVN,
SLV, SOM, SSD, SWZ, SYR,		SWE, TCA, TTO, USA,
TCD, TGO, TJK, TLS, TON,		BHS, FRO, HKG, KOR,
TZA, UGA, UKR, UZB, VNM,		MAC, SXM, SVK, KNA,
VUT, WSM, ZMB, ZWE, LAO,		MAF, VIR
FSM, MDA, STP, PSE, YEM		

Source: World Bank Classification.

TABLE A3
Poisson Regression of Country-level Bilateral Patent Flows by Income Level, 2004–11.
Dependent Variable is Bilateral Trademark Flows from country i to Country j

	(1) All Countries	*(2) Low-to-Low*	*(3) Low-to-Middle*	*(4) Low-to-High*	*(5) Middle-to-Low*	*(6) Middle-to-Middle*	*(7) Middle-to-High*	*(8) High-to-Low*	*(9) High-to-Middle*	*(10) High-to-High*
$\ln(EXP_{ijkt})$	1.196***	0.203	0.788***	0.523***	−0.163***	0.491***	0.983***	−0.186*	0.650***	1.363***
	(0.0445)	(0.181)	(0.125)	(0.0684)	(0.0446)	(0.0970)	(0.0670)	(0.0775)	(0.0427)	(0.0601)
$MARKET_{ijt}$	−0.498*	−2.349***	−0.542	0.443	−0.289	−1.446**	−0.177	−0.432	−0.739	−0.690
	(0.232)	(0.656)	(0.894)	(0.560)	(0.330)	(0.476)	(0.560)	(0.389)	(0.441)	(0.550)
$WEALTH_{ijt}$	−0.707***	3.108***	1.359***	−0.627**	1.339***	−0.313	0.0812	0.971***	−0.971***	−0.248
	(0.0589)	(0.857)	(0.269)	(0.208)	(0.214)	(0.180)	(0.151)	(0.239)	(0.0971)	(0.140)
$\ln(IN\_FDI_{jt})$	0.227***	2.006***	1.760***	−0.382**	0.744***	2.126***	0.201***	2.096***	1.721***	0.0265
	(0.0662)	(0.449)	(0.280)	(0.127)	(0.162)	(0.229)	(0.0564)	(0.190)	(0.139)	(0.0644)
$\ln(OUT\_FDI_{it})$	0.0928	0.246	0.187*	1.589***	0.662***	−0.0692	0.674***	0.595***	0.142**	0.000452
	(0.0489)	(0.191)	(0.0836)	(0.129)	(0.0961)	(0.152)	(0.0766)	(0.0801)	(0.0479)	(0.0532)
IPR_{jt}	9.566***	4.801*	1.313	2.957*	1.466	2.488**	−0.339	1.616*	5.744***	6.748***
	(0.928)	(2.015)	(1.115)	(1.216)	(0.846)	(0.758)	(0.809)	(0.801)	(0.666)	(0.895)
$MARKET_{ijt}$	0.250	0.998	0.0146	0.143	0.146	0.666	0.324	−0.369	0.212	0.388
$\times\ IPR_{jt}$	(0.150)	(0.606)	(0.638)	(0.374)	(0.283)	(0.362)	(0.357)	(0.319)	(0.307)	(0.360)
log	0.478***	−0.812**	−1.742***	1.511***	−1.656***	−0.193	0.267	−1.247***	−0.662***	0.933***
distance	(0.0575)	(0.311)	(0.185)	(0.184)	(0.155)	(0.100)	(0.141)	(0.0941)	(0.0666)	(0.0930)
Border	−0.955***	2.957**	0.586	3.724***	0.829**	1.033***	−2.359***	−0.660	−2.935***	0.171
dummy	(0.203)	(0.953)	(0.310)	(0.540)	(0.294)	(0.156)	(0.303)	(0.486)	(0.371)	(0.312)
Language	−0.145	−0.0285	0.496	1.323***	−1.990***	0.742**	−0.000710	0.288	−0.276	−0.203*
Dummy	(0.108)	(0.698)	(0.610)	(0.172)	(0.441)	(0.249)	(0.146)	(0.193)	(0.299)	(0.0976)
Constant	−45.27***	−33.66***	−29.49***	−37.99***	−1.648	−37.72***	−32.63***	−17.77***	−37.58***	−44.63***
	(1.693)	(5.310)	(3.217)	(1.670)	(1.982)	(1.964)	(1.855)	(1.871)	(1.379)	(1.615)
Observations	59,453	6,374	4,869	7,062	5,639	4,494	6,521	8,201	6,913	9,380

Notes:
(i) All estimations use year fixed effects.
(ii) Robust standard errors in parentheses clustered by bilateral pairs, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

TABLE A4
IV Regression of Bilateral Patent Flows for OECD Countries, 2004–11.
Dependent Variable is Bilateral Trademark Flows from Country i to Country j

	(1) All NICE	*(2) Chemicals*	*(3) Metals and Machinery*	*(4) High-tech*	*(5) Textiles*	*(6) Food and Beverage*	*(7) Other Services*	*(8) Other Manufact.*
$\ln(EXP_{ijkt})$	0.937***	1.303***	0.106	0.958***	0.914**	0.532***	0.445***	0.630***
	(0.0278)	(0.0837)	(0.118)	(0.0498)	(0.324)	(0.121)	(0.0220)	(0.0372)
$\ln(FDI_{ijt})$	0.316**	−0.344**	1.081***	0.553***	0.344	0.589***	0.668***	0.756***
	(0.0963)	(0.132)	(0.121)	(0.121)	(0.208)	(0.179)	(0.105)	(0.106)
$MARKET_{ijt}$	−3.579***	−2.901***	−3.064*	−5.052***	−1.984	−2.393*	−2.608*	−3.998***
	(0.948)	(0.850)	(1.230)	(1.037)	(1.035)	(1.023)	(1.027)	(1.135)
$WEALTH_{ijt}$	−0.467*	−0.139	−0.786**	−0.367	−0.503*	−0.736**	−0.521**	−0.565*
	(0.232)	(0.179)	(0.244)	(0.241)	(0.240)	(0.263)	(0.200)	(0.264)
$\ln(FVA_{jkt})$	0.0686	0.0600	0.220***	0.182***	0.167***	0.0317	0.130**	0.114***
	(0.0352)	(0.0660)	(0.0453)	(0.0255)	(0.0500)	(0.0481)	(0.0464)	(0.0317)
$\ln(OVA_{ikt})$	0.256***	0.120	0.410***	0.309***	0.210***	0.275**	0.294***	0.346***
	(0.0568)	(0.0632)	(0.0626)	(0.0410)	(0.0522)	(0.0840)	(0.0529)	(0.0682)
$\ln(IN\_FDI_{jkt})$	−0.163*	0.110	−0.00666	−0.310**	0.0658	−0.169	−0.129***	−0.334*
	(0.0633)	(0.0941)	(0.0563)	(0.0947)	(0.0736)	(0.0872)	(0.0300)	(0.131)
$\ln(OUT\_FDI_{ikt})$	0.0740	0.260***	−0.118*	0.134*	−0.120	−0.208*	0.0316	0.0130
	(0.0537)	(0.0721)	(0.0586)	(0.0619)	(0.0828)	(0.0934)	(0.0381)	(0.0794)
IPR_{jt}	2.396	−2.290	4.722	−0.312	−0.889	3.815	0.483	3.203
	(3.521)	(3.098)	(3.526)	(4.259)	(3.170)	(3.348)	(3.075)	(4.065)
$MARKET_{ijt} \times IPR_{jt}$	2.540***	2.000***	2.149**	3.639***	1.365*	1.675*	1.901**	2.876***
	(0.624)	(0.545)	(0.798)	(0.681)	(0.674)	(0.681)	(0.671)	(0.758)
log distance	1.141***	1.133***	0.983***	1.213***	1.051***	1.173***	1.019***	1.146***
	(0.253)	(0.241)	(0.223)	(0.163)	(0.217)	(0.286)	(0.241)	(0.218)
Border dummy	−0.172	0.0678	0.750	−0.616	0.152	0.486	0.421	0.209
	(0.655)	(0.510)	(0.579)	(0.537)	(0.625)	(0.736)	(0.527)	(0.604)
Language	−0.693*	−0.599	−1.089***	−0.569	−0.701*	−0.715*	−0.906**	−0.920**
Dummy	(0.280)	(0.331)	(0.256)	(0.315)	(0.281)	(0.336)	(0.311)	(0.314)
Constant	−39.23***	−35.62***	−37.32***	−40.87***	−34.18***	−32.86***	−32.18***	−38.55***
	(5.124)	(4.446)	(4.675)	(5.389)	(6.292)	(5.599)	(4.603)	(5.925)
Observations	72,428	8,893	7,073	5,952	6,097	9,837	13,875	20,701

Notes:
(i) All estimations use year fixed effects. Estimation (1) also contains NICE class fixed effects.
(ii) Robust standard errors in parentheses clustered by bilateral pairs, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.