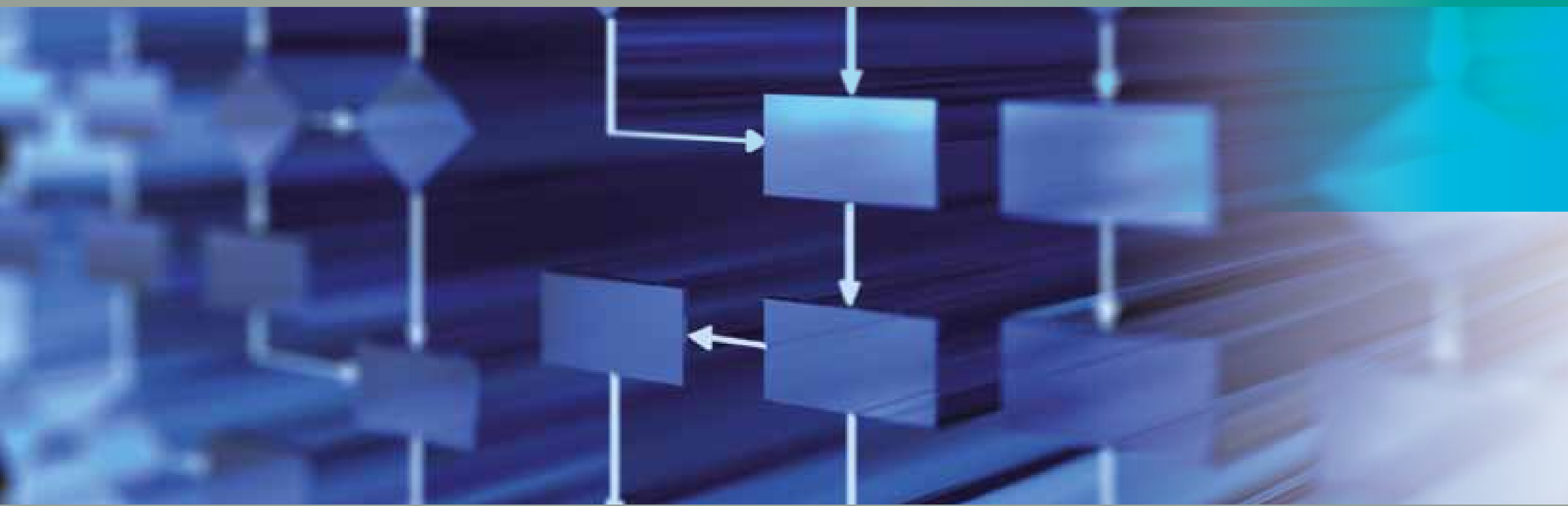

WIPO Economics & Statistics Series

January

2014

Economic Research Working Paper No. 14

An “Algorithmic Links with Probabilities” Concordance for Trademarks For Disaggregated Analysis of Trademark and Economic Data

Travis J. Lybbert
Nikolas Zolas
Prantik Bhattacharyya

WIPO
WORLD
INTELLECTUAL PROPERTY
ORGANIZATION

An "Algorithmic Links with Probabilities" Concordance for Trademarks For Disaggregated Analysis of Trademark and Economic Data

Travis J. Lybbert, Associate Professor, Agricultural & Resource Economics, UC Davis
Nikolas Zolas, Center for Economic Studies, United States Census Bureau
Prantik Bhattacharyya, Software Engineer, Klout

June 2013

Abstract

Trademarks (TMs) shape the competitive landscape of markets for goods and services in all countries through branding and conveying information and quality inherent in products. Yet, researchers are largely unable to conduct rigorous empirical analysis of TMs in the modern economy because TM data and economic activity data are organized differently and cannot be analyzed jointly at the industry or sectoral level. We propose an 'Algorithmic Links with Probabilities' (ALP) approach to match TM data to economic data and enable these data to speak to each other. Specifically, we construct a NICE Class Level concordance that maps TM data into trade and industry categories forward and backward. This concordance allows researchers to analyze differences in TM usage across both economic and TM sectors. In this paper, we apply this ALP concordance for TMs to characterize patterns in TM applications across countries, industries, income levels and more. We also use the concordance to investigate some of the key determinants of international technology transfer by comparing bilateral TM applications and bilateral patent applications. We conclude with a discussion of possible extensions of this work, including deeper indicator-level concordances and further analyses that are possible once TM data are linked with economic activity data.

Keywords: Trademarks, Economics, Concordance, International trade, ISIC, SITC

Acknowledgements: We thank Carsten Fink, Rivka Shenhav, Neil Wilson, and Sacha Wunsch-Vincent for offering feedback and suggestions during early stages of this work.

I. 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 organized differently and can therefore not be analyzed 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.

It is the confluence of two facts that seems largely responsible for the paucity of rigorous empirical research into the relationship between TMs and economic activity. 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 getting TM and economic data to 'speak to each other' 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.

In this paper, 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) or International Standardized Industrial Classification (ISIC). 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 to similarly concord patents to economic data (Lybbert and Zolas, 2012), 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 analyze 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.

II. Background

TM 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%). What is somewhat surprising about this growth is that while overall trademark output has increased dramatically, the level of foreign trademarks (i.e. trademarks applied for in outside jurisdictions), 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 trademark owners to apply for international registrations in countries that have joined this protocol.[1]

Using data from the World Intellectual Property Organization (WIPO), we provide additional perspectives on these trends. We focus mainly on foreign TM applications (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 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 normalize 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 million 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 four times more registrations per $1 million of exports. The pattern of TM use intensity from high income countries is quite distinct: High income countries on average registered 58 foreign TMs in other high income countries but roughly 200 in middle and low income countries for every $1 million 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.

[1] This Protocol materialized from the original Madrid Agreement, which first entered into force in 1892 as a means for international trademark 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 trademark holders to extend the jurisdiction of their trademark to anyone of these countries at any time during the life of the trademark

A. Existing Empirical Research in the Economics of Trademarks

Trademarks are used to differentiate between goods and services offered by competitors within a particular industry. The trademark is intended to reveal information to the consumer regarding both the quality and consistency of a line of goods and services (Landes and Posner 1987; Economides 1987).

For trademark holders, trademarks provide the ability to bypass retailers and communicate directly with customers, along with the flexibility to expand into other product lines and license the trademark to third parties. The economic interpretation of trademarks and why they are important stem from the inherent value in promoting market efficiency and market power while reducing rent-seeking behavior (Ramello 2006), information asymmetry (Economides 1987) and search costs (Landes and Posner 1987). These intangible components make it difficult to assign economic values to trademarks, thus the scope of trademark use in economic studies has been somewhat limited.

The use of trademarks and how it allows the firm to establish and build a particular brand has been rigorously studied within business under “brand management”. In economic studies, trademarks have most widely been used in micro-level studies as a proxy for innovation (Malmberg 2005; Schmoch 2003; Mendonca 2004; Greenhalgh and Rogers 2007; Millet 2009), but also in distinguishing the usage of trademarks across firm size (Allegrazza and Guard-Rauchs 1999; Greenhalgh et al. 2001; Mainwaring et al. 2004) and industry (Greenhalgh et al. 2001; Mainwaring et al. 2004, Schmoch 2003; Jensen and Webster 2004; Loundes and Rogers 2003; Scherer 1983). These findings can be summarized to say that trademarks 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 trademarks can be used to analyze certain relevant aspects of innovation and industrial change. They encourage greater studies that use trademark data and explain how trademark-based indicators can provide a partial measure of innovative firm output, international patterns of specialization, links between technology and marketing, as well as the evolution of firm organization and structure. Regarding firm size, the use of trademarks is inconclusive as one study shows that trademark usage increases with firm size (Allegrazza and Guard-Racuhs 1999), while another shows the opposite effect (Greenhalgh et al. 2001). In a more recent study, Mainwaring et al. (2004) show an inverted U-shape relationship with regards to firm size and trademark activity.

There are a limited number of papers using aggregate measures of trademarks to study a wide range of economic topics. One paper looks at country-level differences in usage (Baroncelli et al. 2005) and finds that rich countries dominate trademark activity and that trademarks provide information on the global distribution of intellectual property rights (IPRs) and evidence of investment in reputational assets. Other papers look at how trademarks can infer economic growth (Yorukogly 2000), trade specialization (Fink et al. 2003; Mangani 2007) and be used as a form of protectionism (Baroncelli et al. 2004). Fink et al. (2003) use aggregate trademarks to infer both the variety and quality of trade flows and then use it to test the Linder hypothesis. Mangani (2007) uses aggregate trademarks in a similar manner to infer the number of varieties across trademark classes (extensive margin), as well as the number of varieties within trademark classes (intensive margin). The author then uses these two measures to infer the quality of a country’s goods and services, assuming that higher quality goods are trademarked across a higher number of classes. These two studies are good examples of how trademarks can be used in future empirical trade studies to benchmark quality and estimate varieties.

The use of trademarks in economic studies has been limited because of the difficulty in assigning economic values to trademarks and problems with aggregation since trademarks 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 trademark activity, as well as decompose the use of trademarks across many different sectors. Economists will better understand how trademarks fit into the overall innovation chain, as well as estimate the value of trademarks from their different uses in a variety of industries. Matching trademark data with trade flows will also provide information regarding the exporting behavior of firms, quality within and across varieties and intellectual property rights, since trademarks 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 International Standardized Industrial Classification (ISIC) system and Standardized International Trade Classification (SITC) system, respectively. A conventional concordance approach would link a classification level in NICE to a comparable classification level in ISIC or SITC.

Unfortunately, such a conventional approach is complicated by the fact that the NICE system is structured very differently than SITC or ISIC. The SITC and ISIC systems are designed to facilitate the collection and processing of data and therefore have an explicit multi-leveled 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 pre-defined 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 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 is not organized by GS indicators. One possible remedy to this problem would be to directly classify each TM registration according to SITC or ISIC. This approach would generate a supplementary data file for any given TM database that contains a list SITC or ISIC 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 and ISIC – and consequently each NICE class potentially maps to multiple ISIC or SITC 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 SITC or ISIC 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 U.S. 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 SITC or ISIC is likely to be noisier outside than inside the U.S.

III. 'Algorithmic Links with Probabilities' Matching: 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 Algorithmic Links with Probabilities (ALP) matching. This approach has been used elsewhere to match patent data to economic activity (Lybbert and Zolas,2012), 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.

ALP 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 in order to minimize 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 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 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 databases 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.

---

[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
[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. A description of the comparison between the ALP concordance constructed using USPTO data versus ROMARIN data is described in the Appendix section I.

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 generalized in Figure 1, 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 ISIC or SITC classification systems[5]. Often times, these descriptions will contain product types, uses and one or two sentences with a brief description of the industry. For our purposes, we utilize the hierarchical structure of each industry classification system and utilize the most disaggregate descriptor available, which will consist of either 4-digit ISIC or 4-5 digit SITC codes.

A.1 Generating Keywords for Industry Descriptors

To generate the keywords used in the matching process, we do multiple things. 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", "product" and more. We also remove the filler words, such as "the", "as", etc. so that the remaining list of terms is 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 specialized 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 4 or 5-digit level industry will have between one and dozens of keywords associated with it. These are then queried and matched with the TM keywords.

A.2 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 way[7]. After comparing all these options, we settled on a relatively simple approach that converts each GS indicator phrase – whether pre-defined or user-defined – into a batch of keywords and expands these keyword batches to include their plural / singular analogs.

[5] For instance, the latest publications for both the ISIC Rev. 4 and SITC Rev. 4 can be found here: http://unstats.un.org/unsd/cr/registry/regdnld.asp

[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.

For each of the TMs in our data, we extract the text associated with the GS indicator(s). Multiple indicators are separated by semi-color (;). 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 a large density of key-phrases in the indicator text (more than 70%), we run the text through a text-to-keyword extraction software (Topia TermExtract 1.1) 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%) The end result of this process is a rich set of TM keywords in batches that correspond to the semi-colon delimited GS indicators as chosen by the applicant.

A.3 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 utilize "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 4 or 5-digit industry code. We do not require 100% 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 or 5-digit industry codes (of which there exist more than 400 ISIC industries and more than 900 SITC industries) 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 nonsensical 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 cutoff threshold for aggregated weights (2%). 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%, 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 or 2-digit SITC). The final result is that each NICE code has the potential to match up with roughly 100 SITC and 50 ISIC codes on average, rather than 900 SITC and 400 ISIC codes in the initial stage. This provides us with cleaner frequencies and minimizes 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 each industry mapping to anywhere between one and dozens of NICE classes. To further reduce potential errors and/or biases introduced by the matching process, we reweight the results according to the weighting scheme utilized in Lybbert & Zolas (2012).

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 trademark class *j* and B_i be the *ex ante* probability of a trademark being matched with industry class *i*. Assuming *J* is the total number of different trademark classes available (45), Bayes rule gives the *ex post* probability of A_j conditional on observing B_i where:

$$\Pr\left(A_j \middle| B_i\right) = \frac{\Pr\left(B_i \middle| A_j\right)\Pr(A_j)}{\Pr(B_i|A_1)\Pr(A_1) + \ \dots \ + \Pr\left(B_i \middle| A_J\right)\Pr\left(A_J\right)} \tag{1}$$

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

$$W_{ij}^{H} = \frac{\Pr\left(B_i \middle| A_j\right)\left(\Pr\left(A_j \middle| B_i\right)/J\right)}{\Pr(B_i|A_1)\left(\Pr\left(A_j \middle| B_i\right)/J\right) + \ \dots \ + \Pr\left(B_i \middle| A_J\right)\left(\Pr\left(A_j \middle| B_i\right)/J\right)} \tag{2}$$

As Lybbert & Zolas (2012) note, this weighting scheme prioritizes (i.e. gives higher weights) to the most frequent matches for very specifically defined industries/trademarks, while giving less weight to the broadly defined industries/trademarks who might have a large number of erroneous matches due to the nature of their definition.

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

Once the full process has been completed, we find that each NICE class maps onto roughly 8-10 four-digit industries on average, while each industry maps onto approximately 5 NICE classes on average. This completes the construction of the concordance.

IV. Using ALP Concordances to Jointly Analyze TMs & Economic Data

With the concordance, it is now possible to jointly analyze industry-level economic activity with trademarks which will allow researchers to better understand the value of trademarking and branding, how industry life-cycles are influenced by trademarking and much more.

A. Trademark Intensity by Income Group

As a first exercise, we start by analyzing country-level differences in trademarking by industry. Specifically, we look at the trademarking intensity by industry across countries of different income levels in order to identify patterns of trademark growth and specialization. For data, we use the WIPO IP Statistics website, which contains trademark output by NICE class for 192 countries between 2004 and 2008. A summary chart in Figure 3 shows the growth of trademarks in Goods (NICE classes 1 through 34) and Services (NICE classes 35 through 45) for three separate 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 5-year window with trademark output in both goods and services increasing by around 50% in High income countries. Middle income countries experienced even more rapid growth in Services, where trademark output more than doubled in the time period. Meanwhile, Low income countries experienced little-to-no growth throughout.

To derive our measure of intensity, we proxy for productive output by using total export value, since other industry-level data is unavailable for this many countries[9]. The export data is initially organized by two-digit SITC Rev. 2 and was gathered from UN COMTRADE database for the years 2004-2008. 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 features from this figure is that Middle income countries are the most trademark intensive in nearly every trademark class and are more than twice as intensive as High income countries across all Goods. 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 (China applied for more than 150,000 trademarks in Pharmaceuticals between 2004 and 2008 compared to roughly 50,000 trademarks applied for by US firms in the same time period), countries such as Russia and Mexico are also very active with trademarking. For Low income countries, they are the most active in "Yarns and Threads" (Class 23) and "Alcoholic Beverages" (Class 33) relative to High income countries, with more than 4 times the intensity in each of these classes.

B. Intensity of Foreign Trademark Transfers by Income Group

In the next exercise, we look at the intensity of different types of trademark classes being transferred between income groups, which was similar to the earlier country-level analysis done in Table 1 and Figure 1. However, in this case, we break down the analysis by industry-type. We again use total export value as our measure for relative intensity and focus our attention on a few key trademark 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 Class 9, 38 and 42 to

[8] Note that the World Bank classifies countries according to 4 income groups: High, High Middle, Low Middle and Low. We combined Low and Low-Middle income countries to form one 1 Low Income group.
[9] We do have industry-level Value Added and Production for OECD countries, which we look at later in the paper.

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 trademark intensity is highest when the High income country is the origin. What is more interesting is that the intensity is often highest for transfers to low income nations (especial ly for “High-tech” products) and lowest for transfers to high income nations. If we were to conduct the same exercise for foreign patents[10] using the ALP Patent Concordance (see Lybbert & Zolas 2012), then this finding differs greatly from foreign patenting intensity (see Figure 6), where intensity is highest to high income nations and virtually nonexistent to low income nations. This seems to be a clear indication that countries incorporate different strategies for their intellectual property when operating abroad. For new technologies (proxied for by patents), countries tend to worry most about other advanced nations being able to reproduce or replicate that specific technology without worrying about low income nations. On the other hand, for finished products that are ready to come to market, countries apply for trademarks more generally across all income groups, with more emphasis on the low income groups. This may have to do with the prevention of counterfeit goods in low income nations or possibly because trademarks are cheaper and easier to apply for than patents. Regardless, this type of analysis is possible using industry-level concordances for trademarks and patents.

C. Trademark Intensity for OECD Nations

Our next exercise looks at trademark intensity as a proportion of a country’s value-added. This data is broken down by industry in the OECD Structural Analysis (STAN) database. The value-added data[11] is organized by 2-4 digit ISIC Rev. 3, which is mapped into the NICE classification using the ALP concordance. We first look at the intensities for all 45 NICE classes for the OECD in whole, and then separate the OECD countries into Low, Medium and High TM intensity countries based on their rankings domestic TM output cutoff conditions[12]. These figures can be found in Figure 7.

Across all of the OECD countries, we can see that most trademarking activity is relatively consistent across all industries, with domestic TM intensities in the range of 50-100 TM’s per $billion in value-added. This pattern holds between both Goods (Class 1-34) and Services (35-45). This consistency is also found in exported and imported TM’s, with roughly similar ratios of intensity. Amongst Goods, certain industries do stand out in terms of intensity. Class 15 (“Musical Instruments”) has a relatively high TM intensity compared to other industries, and interestingly, textiles (Class 23, 24 and 25) all have much higher TM intensities in trade (exports and imports) relative to domestic TM output. In Services, we find that Class 38 (“Telecommunications”) has a much lower domestic TM intensity, with very few observations of tradable TM intensities. Amongst the tradable intensities, Class 43 (“Services for providing food; temporary accommodation”), has a much lower tradable TM intensity, while classes 42 (“High-tech Services”) and 44 (“Medical services”) both have much higher tradable TM intensity relative to domestic intensity. These same patterns hold and are consistent across the Low, Medium and High TM intensity countries.

[10] We foreign patent data from WIPO’s IP Statistics for the same years (2004 to 2008). The patents are initially classified using the International Patent Classification (IPC) system, which differs from the NICE classification system. We first utilize the ALP Patent Concordance to map the patents into SITC Rev. 2. We then layer the ALP Trademark Concordance on to then map from the patents from the SITC to NICE classification system.

[11] We use the “Value Added (at Current Prices)”from the STAN Database

[12] We looked at the relative TM intensity of each OECD country to all OECD countries, meaning that we divided the domestic TM intensity by the weighted average intensity of all OECD countries. The “Low” intensity countries had an average intensity that was less than 20 times the OECD average and consists of 10 countries (see Appendix), while the “Medium” intensity had domestic TM intensities that were between 30 and 100 times the OECD average and consists of 11 countries. Finally, the “High” intensity countries had domestic TM intensities of 100 times or greater. Note that the relative levels are so high mainly due to the extremely low TM intensities (their combined intensity is less than 1/5 of the OECD average) of Korea and Japan, whose combined weight make up a significant share of the OECD value-added. However, the rankings of the countries still persist and we thought it helpful to break them up by intensity in order to find patterns in the data.

We next take our descriptive analysis a step further and analyze precisely how different country and country-TM class factors influence the decision to trademark in a jurisdiction. To do this, we utilize similar datasets and measure the propensity to trademark in a jurisdiction based on trade flows, patent flows, FDI inflows and outflows, income and other variables.

D. Regression Analysis of Foreign Trademark Intensity

ALP concordances for TM and economic data open new possibilities for more rigorous empirical analysis. In this section, we look at the determinants of foreign trademark 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 4 and 5 highlighted some interesting patterns and differences in foreign trademarking and patenting behavior for three types of industries and the intention of this analysis is to shed additional light as to what could be causing these differences. This next exercise wants to look for further differences in the use of intellectual property abroad by comparing the determinants of foreign patenting and trademarking.

Since trademarks are traditionally used for “branding” and are related to the sale of final goods, it makes sense for bilateral foreign trademark flows to be related closely to bilateral trade flows. For this reason, our key regressor will be trade, followed by other measures of country-industry output and more.
Our main specification is:

$$\begin{aligned} TM_{ijkt} = {} & \alpha_0 + \alpha_1 \ln\left(TRADE_{ijkt}\right) + \alpha_2 \ln(FDI_{ijt}) + \alpha_3 MARKET_{ijt} + \alpha_4 WEALTH_{jt} + \alpha_5 \ln(OVA_{ikt}) \\ & + \alpha_6 \ln\left(FVA_{jkt}\right) + \alpha_7 \ln(IN\_FDI_{ikt}) + \alpha_8 \ln(OUT\_FDI_{jkt}) + \alpha_9 \ln(IPR_{jt}) + X_{ijt} + \delta_t + \delta_i \\ & + \delta_j + \varepsilon_{ijkt} \end{aligned} \tag{3}$$

where *i* is the origin country, *j* is the destination country, *k* is the NICE class group and *t* is the year. *TM* measures bilateral exported trademarks, *TRADE* is bilateral trade flows, *FDI* is bilateral FDI flows per capita (measured at the country-level since industry-level data is not available), *MARKET* is relative market size (measured as destination country’s GDP divided by the origin country’s GDP), *WEALTH* is the relative wealth (measured as destination country’s GDP per capita divided by the origin country’s GDP per capita), *IN_FDI* and *OUT_FDI* are the destination country’s FDI inflows and origin country’s FDI outflows, *OVA* is origin country value added and *FVA* is the destination country value-added, and *X* are trade costs such as distance, contiguity and sharing the same language. We also include time and country fixed effects. The list of sources of data and how they are organized and converted to the NICE group are listed in Table 2, along with summary statistics.

We run the same exact specification for patents and break down the analysis by industry types to highlight specific differences in intellectual property by industry. Since we are measuring trademark and patent flows, which are discrete variables, we need to choose an estimator for count data. Therefore, we use the Poisson-based estimator in our analysis. 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 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 Class 1, 2, 3, 4 & 5), “Metals & Machinery” (NICE Class 6- 8, 12 & 14) , “High-Tech” (NICE Class 9, 38 & 42), “Textiles” (NICE Class 22- 26), “Food & Beverages” (NICE Class 27-34), “Other Manufacturing” (NICE Class 10, 11, 15-21, 27, 28) and “Other Services” (NICE Class 35-37, 39-41, 43-45).

We run the analysis across all OECD countries between 2004 and 2008. We use OECD data since industry-level data is most readily available for this set of countries. Since our sample runs from 2004 to 2008 for 31 different countries, the total number of possible observations in this sample is 209,250. However, due to numerous instances of missing data (particularly for FDI flows) and cases where there was no variation for bilateral country pairs (i.e. observations dropped out), our estimating sample consists of roughly 60,000 observations. To ensure consistency between the patent and trademark datasets, we drop the observations that have missing patent or missing trademark 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 trade plays a very important role in the transfer of intellectual property abroad, with exports being highly significant and positive for both trademarks and patents. Across all NICE classes, both patents and trademarks appear to be equally sensitive. However, certain types of goods are more sensitive to trade flows than others. For trademarks, the elasticity is mostly similar across all types of goods, with "High-tech" goods being the most sensitive to trade, with nearly twice the elasticity of the other goods. For patents, 'High-tech' goods are more than 3 times as sensitive as the other goods. We also see that 'Chemicals' have a much higher elasticity than other goods. Among the least elastic goods, it appears that "Food & Beverages" are the least sensitive.

In other variables, we also see some slight differences in the behavior of industries who trademark and patent. We find that FDI flows are a positive and significant determinant of trademark flows, but have no effect for patents. This holds true across nearly all the types of goods. We also find that the relative wealth of the destination country (in relation to the origin country) has a negative effect for trademarks, implying that firms are more likely to trademark in poorer countries. This is especially true for "Textiles", which has a high negative elasticity from relative wealth. On the other hand, the relative wealth has no effect on patents, while relative market size is important.

In terms of value-added, we find that the destination country's value-added has a positive and significant impact on both trademarks and patents, with subtle differences across industries. Value-added appears to have the strongest effect in "Textiles" relative to other goods. This is especially true for patents. We also find that the Aggregate industry measures of FDI inflows and outflows have a positive and significant impact for trademarks, with little-to-no effect for patents. This is interesting and somewhat surprising since we would expect the opposite to occur since trademarks typically operate for final goods, which would not necessitate investment or the purchase of subsidiaries abroad. We will investigate this effect later on in the paper. More interesting though, is the fact that a destination country's IPR environment has a negative effect on trademarks, but a very strong and positive effect on patents. Among the gravity variables (distance, border effects and language dummies), all of the signs and significance point in the direction we expected for trademarks, with distance having a negative impact and both the border dummy and language dummy having a positive effect. However, this does not hold true for patents, where we find varying effects of distance for different types of goods. We see that distance seems to have a negative effect on "Food" and "Textiles", both of which are relatively low-technology goods. On the other hand, distance appears to have a positive effect on all the other goods. This seems to imply that perhaps firms are more interested in coverage when it comes to applying for patents abroad.

The consensus from this analysis points to several subtle differences in the behavior of industries who transfer their intellectual property abroad in the form of patents and trademarks. We have shown that across all industries, patents and trademarks appear to be equally sensitive to trade flows, with similar patterns of behaviors for each type of industry (namely, that "High-tech" goods are the most sensitive to trade flows). We also find that FDI, both in terms of flows and in aggregate inflows and outflows, appears to be a good predictor for trademarks, but not so much for patents. Finally, intellectual property rights (IPR) have a very strong and significant effect on the decision to patent abroad, but a negative effect on the decision to trademark. This is also very interesting since both require a strong IPR environment in order for the trademark or patent to be effective.

D.1 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.

As an additional check, we run an instrument 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 5, 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 alter the outcome of trade elasticity too much for either patents or trademarks. However, our coefficient for FDI flows turn negative and significant. This removes some of the doubt caused earlier by the fact that trademarks were positively impacted by FDI flows, while patents had no effect. On the other hand, it does raise questions as to why this coefficient suddenly became negative and significant. Outside of these coefficients, we find very few changes in the magnitude and signs of the other variables.

To summarize, we have explored differences in behavior of firms taking out intellectual property abroad using count-estimation and an IV approach. We have identified several industry-specific differences, such as certain goods being more/less sensitive to exports and relative market size and wealth. Among the key differences between trademarking and patenting behavior is that patents rely heavily on the destination country's IPR environment, while trademarks consider this to be unimportant. We also find that trademarks are significantly influenced by both FDI flows and aggregate FDI inflows/outflows from the origin and destination country.

This exercise was done simply to illustrate some potential uses of a NICE class-industry concordance to enhance our understanding of technology transfer and more. More detailed analysis needs to be done to investigate the causes for these differences, which we leave up to future researchers.

VII. Conclusions & Future Work

Although trademarks 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 for assessing this importance at the economy-level because TM data and economic activity data are organized differently and cannot be analyzed 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 that previously possible.

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

To build this linkage, we develop an algorithmic approach we call ‘Algorithmic Links with Probabilities’ (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 trademark flows can now be organized by SITC and ISIC classification systems, we can merge other datasets organized by these classification systems, such as trade elasticities of substitution from Broda and Weinstein (2006) and the Rauch Classification of Goods (Rauch 1999).

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 modeling 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 present 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 analyzed 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 Aplique, Vol. 52, No. 2, pp. 51-68.

Baroncelli, Eugenia and Carsten Fink and Beata Javorcik. (2005) “The Global Distribution of Trademarks: Some Stylized Facts.” The World Economy, Vol. 28, No. 6, 765-782.

Baroncelli, Eugenia and Ekaterina Krivonos and Marcelo Olarreaga. (2004) “Trademark Protection or Protectionism?” World Bank Policy Research Working Paper.

Broda, Christian and David Weinstein. (2006) “Globalization and the Gains from Variety” Quarterly Journal of Economics, Vol. 121, No.2, 541-585.

Economides, Nicholas. (1987) “The Economics of Trademarks” Trademark Register, Vol. 78, 523-539.
Fink, Carsten and Beata Javorcik and Mariana Spatareanu. (2003) “Income-Related Biases in International Trade: What do Trademark Registration Data Tell Us?” World Bank Policy Research Working Paper.

Greenhalgh, C. and 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.

Greenhalgh, Christine and Mark Rogers. (2007) “Trade Marks and Performance in UK Firms: Evidence of Schumpeterian Competition through Innovation.” Working Paper.

Jensen, Paul and Elizabeth Webster. (2004) “Patterns of Trademarking Activity in Australia.” Melbourne Institute Working Paper No. 2/04.

Landes, William and Richard Posner. (1987) “Trademark Law: An Economic Perspective.” Journal of Law and Economics, Vol. 30, No. 2, pp. 265-309.

Loundes, Joanne and Mark Rogers (2003) “The Rise of Trade Marking in Australia in the 1990s”, Melbourne Institute Working Paper No. 2/04.

Mainwaring, L. and N. Moore and P. Murphy. (2004) “Trademark Holdings of Production Firms in Britain and Ireland.” Working Paper.

Malmberg, Claes. (2005) “Trademarks Statistics as Innovation Indicator? – A Micro Study.” Centre for Innovation, Research and Competence in the Learning Economy Working Paper.

Mangani, Andrea. (2007) “Measuring Variety and Quality of Products with Trademarks.” International Economic Journal, Vol. 21, No. 4, 613-631.

Mendonca, Sandro and Tiago Pereira and Manuel Godinho. (2004) “Trademarks as an Indicator of Innovation and Industrial Change.” LEM Working Paper.

Millot, Valentine. (2009) “Trademarks as an Indicator of Product and Marketing Innovations.” OECD Science, Technology and Industry Working Paper.

Ramello, Giovanni. (2006) “What’s in a Sign? Trademark Law and Economic Theory.” Working Paper.

Rauch, James. (1999) “Networks Versus Markets in International Trade.” Journal of International Trade, Vol. 48, No. 1, 7-35.

Rujas, Javier. (1999) “Trademarks: Complementary to Patents.” World Patent Information, Vol. 21, No. 1, 35-39.

Santos Silva, J.M.C, and Silvana Tenreyro. (2006) “The Log of Gravity.” The Review of Economics and Statistics, Vol. 88, No. 4, 641-658.

Scherer, F.M. (1983) “The Propensity of Patent.” International Journal of Industrial Organization, Vol. 1, No. 1, 107-128.

Schmoch, Ulrich. (2003) “Services Marks as Novel Innovation Indicator.” Research Evaluation, Vol. 12, No. 2, 149-156.

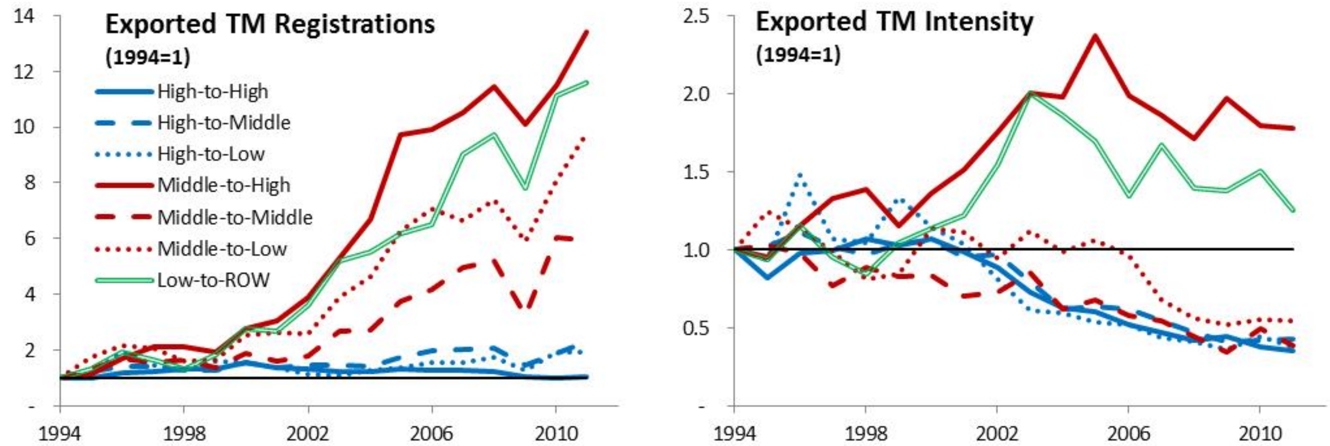

Figure 1: Total exported TM registrations (left) and TM intensity for exported registrations (right) from and to different income classes (ROW=Rest of World)

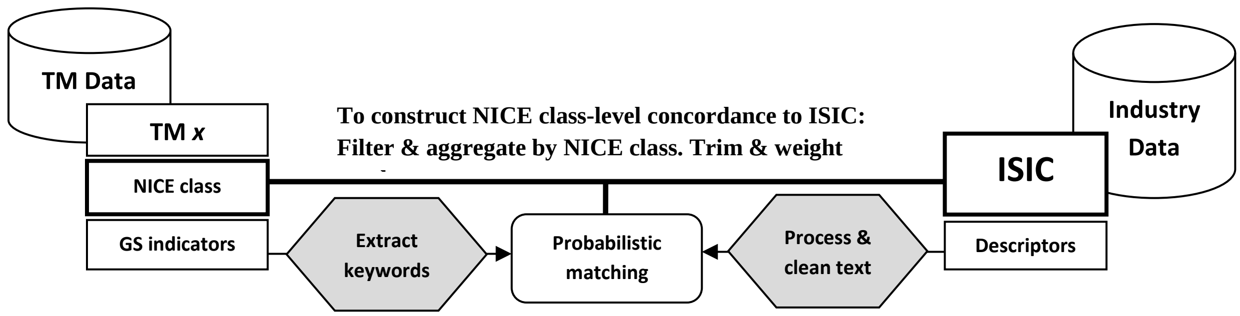

Figure 2: Heuristic depiction of ALP matching process for constructing NICE class-level concordances

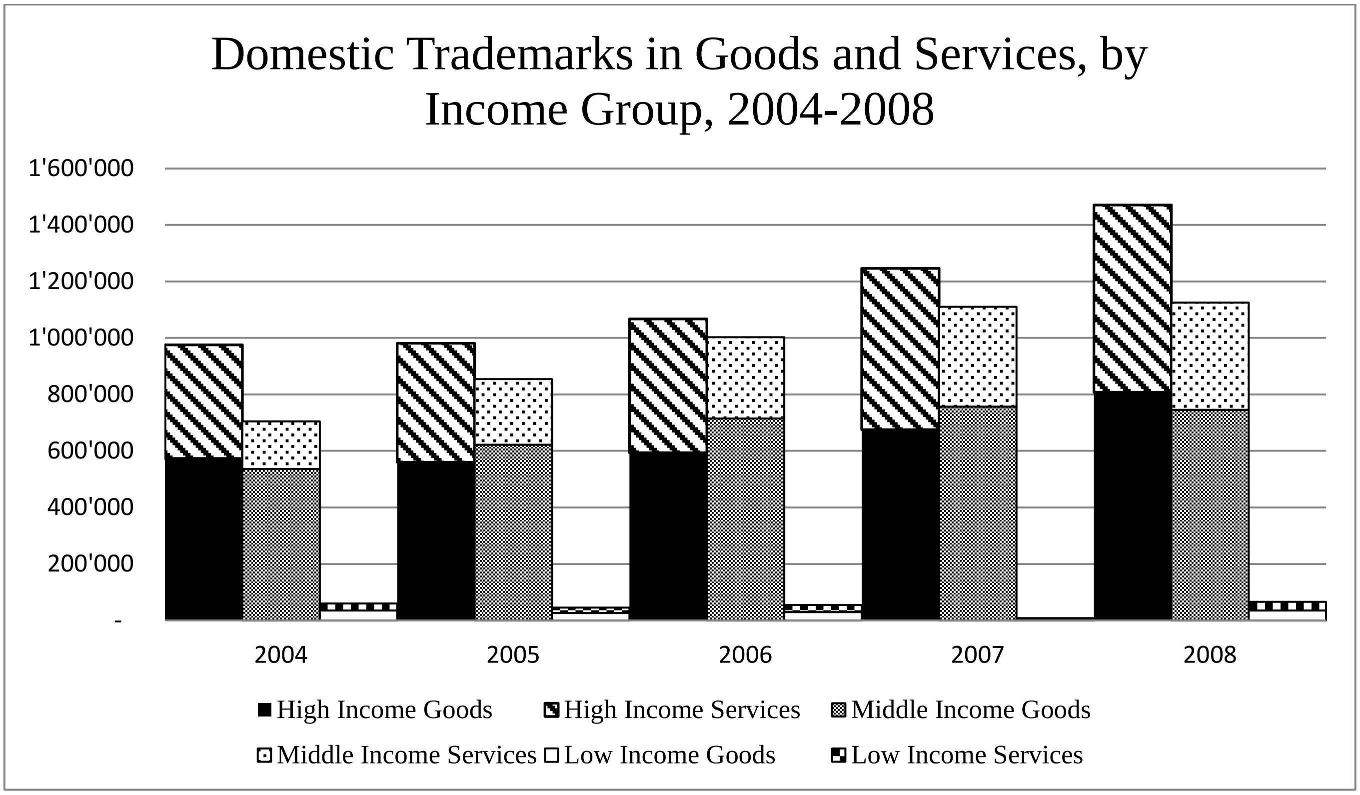

Figure 3: Trademark Output in Goods (NICE Class 1-34) and Services (NICE Class 35-45) by Income Group, 2004-2008

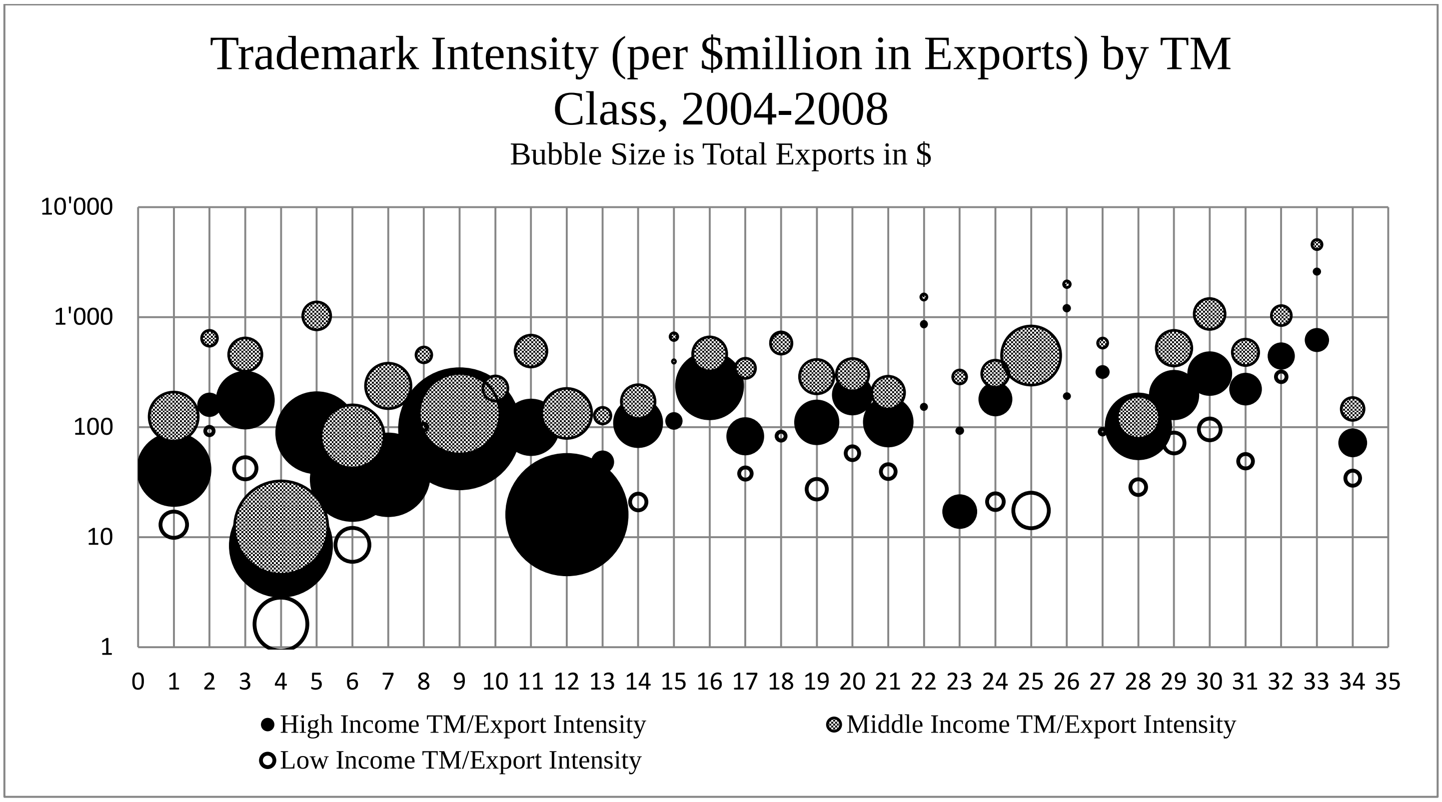

Figure 4: Trademark Intensity (per $ million in Exports) for All Goods by Income Group, 2004-2008

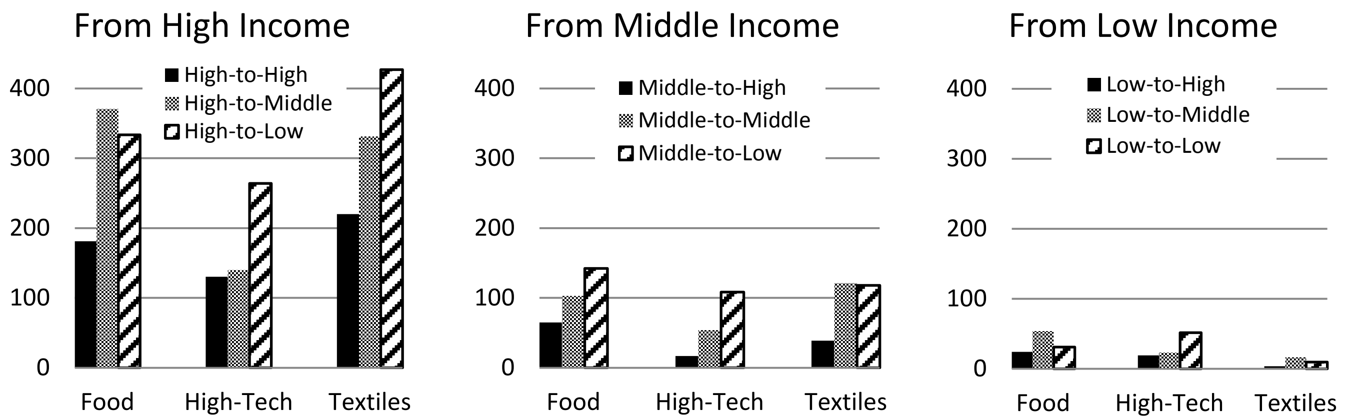

Figure 5: Foreign Trademark Intensity (per $ million in Exports) for various Goods and Services by Income Group, 2004-2008. “Food” consists of NICE Classes 29, 30 and 31. “High-Tech” consists of NICE Classes 9 and 42. “Textiles” consists of NICE Classes 23, 24 and 25.

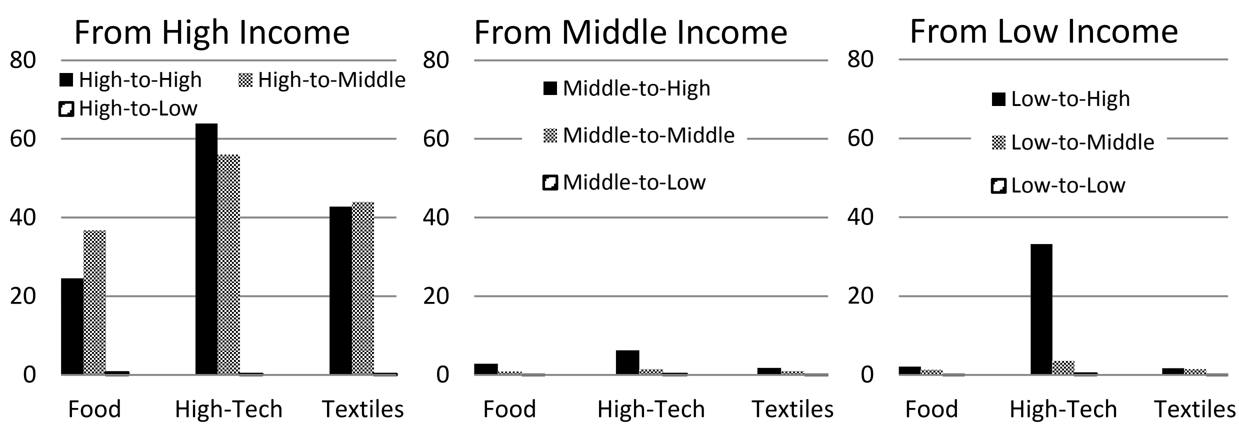

Figure 6: Foreign Patent Intensity (per $ million in Exports) for various Goods and Services by Income Group, 2004-2008. “Food” consists of NICE Classes 29, 30 and 31. “High-Tech” consists of NICE Classes 9 and 42. “Textiles” consists of NICE Classes 23, 24 and 25.

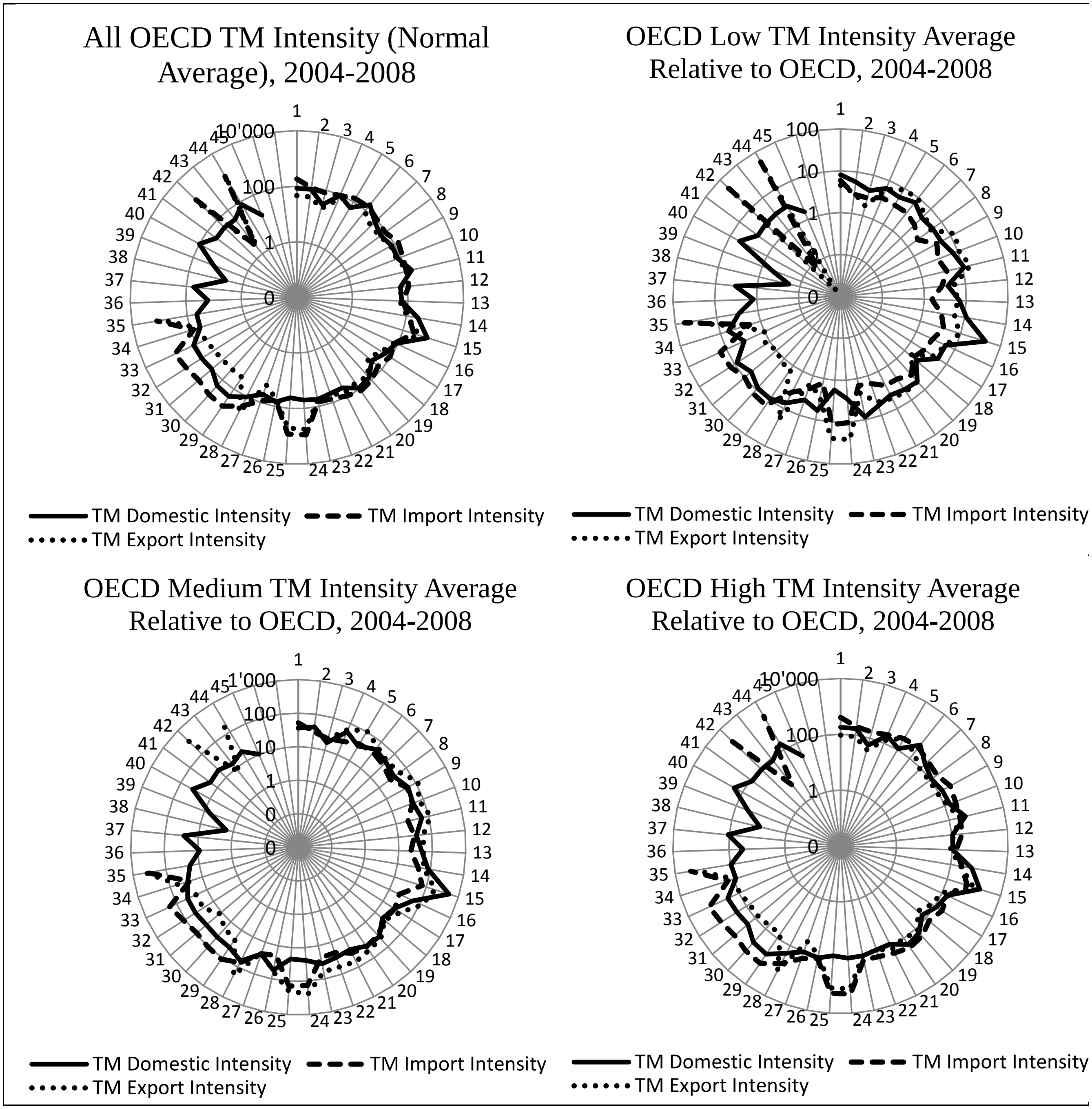

Figure 7: Trademark Intensity Levels for OECD Countries.

Table 1: Average (1994-2011) exported TM registrations and TM intensity from and to different country income classes

Average Exported TM Registrations		From						
		High	*% Total*	**Middle**	*% Total*	**Low**	*% Total*	**Total**
To	**High**	298,600	*56%*	22,229	*44%*	1,060	*31%*	321,889
	Middle	180,381	*34%*	19,578	*39%*	1,496	*44%*	201,455
	Low	51,719	*10%*	8,677	*17%*	881	*26%*	61,277
	Total	530,700		50,484		3,437		
Average Exported TM Intensity		**From**						**Weighted Average**
		High	*% Own*	**Middle**	*% Own*	**Low**	*% Own*	
To	**High**	58	*100%*	17	*20%*	5	*23%*	55
	Middle	206	*353%*	85	*100%*	29	*135%*	193
	Low	197	*337%*	71	*83%*	22	*100%*	177
Weighted Average		122		52		20		

Table 2: Data Sources for 2004-2008

Type of Data	Definition:	Source:	Original Classification:	Mean/ S.D.	# of Observations
Bilateral TM Flows	Number of exported TM applications	WIPO	NICE Class	12.77957 (44.9776)	209,250
Bilateral Patent Flows	Number of exported patent applications	WIPO	4-digit IPC, converted to 4-digit ISIC Rev. 3	6.178613 (93.00802)	208,950
Bilateral Trade Flows	Export values ($)	UN Comtrade	4-digit SITC Rev. 2	141513.8 (872886.3)	206,155
Bilateral FDI Flows	Amount of FDI exported ($)	OECD	Country Level	1446.348 (7267.496)	146,070
Country-Industry FDI	Aggregate FDI inflow & outflow by industry ($)	OECD	2,3,4-digit ISIC Rev. 3	310517.1 (561374.2)	202,500
Country GDP	Aggregate GDP ($)	World Bank	Country Level	1.22E12 (2.43E12)	209,250
Country-Industry VA	Aggregate Value Added by Industry ($)	OECD	2,3,4-digit ISIC Rev. 3	9.48E11 (7.06e+12)	185490
Country IPR	Country Intellectual Property Rights (1-5)	Ginartes & Park (2008)	No conversion	4.373977 (.2870399)	195,750
Gravity variables	Distance (km), Border, Language	CEPII	No Conversion		

All Industry-level data was mapped to NICE Classes using the 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 Bilateral Trademark flows for OECD Countries, 2004 – 2008. Dependent Variable is Bilateral Trademark Flows from country *i* to country *j*.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All NICE	Chemicals	Metals & Machinery	High-Tech	Textiles	Food & Beverage	Other Services	Other Manufact.
$\ln(TRADE_{ijkt})$	0.245***	0.243***	0.353***	0.466***	0.215***	0.290***	0.160***	0.348***
	(0.00591)	(0.0134)	(0.0155)	(0.0170)	(0.0269)	(0.0117)	(0.00823)	(0.0189)
$\ln(FDI_{ijt})$	0.0741***	0.0631***	0.0385***	0.0474***	0.0533*	0.0586***	0.106***	0.0507***
	(0.00394)	(0.00998)	(0.00995)	(0.0111)	(0.0237)	(0.00827)	(0.00779)	(0.00856)
$MARKET_{ijt}$	-0.00299***	-0.00122	-0.00577***	-0.00169**	-0.00313	-0.000778	-0.00424***	-0.00252***
	(0.000448)	(0.00136)	(0.00148)	(0.000584)	(0.00326)	(0.000558)	(0.000950)	(0.000691)
$WEALTH_{ijt}$	-0.170***	-0.180***	-0.0484	-0.270***	-0.442***	-0.0478	-0.204***	-0.129***
	(0.0143)	(0.0426)	(0.0329)	(0.0488)	(0.124)	(0.0293)	(0.0289)	(0.0238)
$\ln(FVA_{jkt})$	0.0792***	0.101***	-0.00503	0.0900**	0.300***	0.165***	0.153***	0.0926**
	(0.00859)	(0.0258)	(0.0173)	(0.0294)	(0.0493)	(0.0171)	(0.0177)	(0.0317)
$\ln(OVA_{ikt})$	0.0695***	0.570***	-0.0970***	-0.0581	-0.395***	0.0238	0.198***	-0.0195
	(0.0124)	(0.0309)	(0.0202)	(0.0337)	(0.0442)	(0.0186)	(0.0170)	(0.0372)
$\ln(IN\_FDI_{ikt})$	0.00185***	0.00804	0.00595	-0.000892	0.459***	0.0311*	0.000180	-0.00320
	(0.000523)	(0.00553)	(0.00587)	(0.00198)	(0.103)	(0.0125)	(0.000465)	(0.00395)
$\ln(OUT\_FDI_{jkt})$	0.00150***	0.0125***	-0.0157*	-0.000465	0.278***	0.0755***	0.00290***	0.0155**
	(0.000302)	(0.00294)	(0.00727)	(0.000587)	(0.0422)	(0.0114)	(0.000403)	(0.00472)
IPR_{jt}	-1.953***	-1.219***	-1.974***	-2.743***	-0.433	-1.354***	-2.459***	-1.858***
	(0.122)	(0.278)	(0.283)	(0.323)	(0.894)	(0.327)	(0.294)	(0.263)
log Distance	-0.257***	-0.197***	-0.166***	-0.0870*	-0.361***	-0.186***	-0.294***	-0.221***
	(0.0101)	(0.0231)	(0.0221)	(0.0416)	(0.0653)	(0.0246)	(0.0189)	(0.0241)
Border Dummy	0.140***	0.127**	0.125**	0.0991	-0.0903	0.152***	0.267***	0.100**
	(0.0171)	(0.0443)	(0.0438)	(0.0612)	(0.127)	(0.0400)	(0.0341)	(0.0354)
Language	0.490***	0.355***	0.313***	0.498***	0.289	0.313***	0.655***	0.419***
Dummy	(0.0229)	(0.0561)	(0.0489)	(0.0550)	(0.171)	(0.0531)	(0.0403)	(0.0545)
Constant	8.640***	-7.566***	13.04***	12.71***	6.515	4.116*	7.818***	9.122***
	(0.691)	(1.427)	(1.445)	(1.668)	(4.509)	(1.659)	(1.473)	(1.391)
Log-likelihood								
Observations	59493	7344	7475	4414	3289	8257	14751	13963
Pseudo R^2	0.818	0.845	0.798	0.872	0.675	0.743	0.802	0.753

Robust Standard errors in parentheses, * p<0.05, ** p<0.01, *** p<0.001. All estimations use year-fixed effects and have individual country effects (origin and destination fixed effects). Estimation (1) also contains NICE class fixed effects.

Table 4: Poisson Regression of Bilateral Patent Flows for OECD Countries, 2004 – 2008. Dependent Variable is Bilateral Patent Flows from country *i* to country *j*.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All NICE	Chemicals	Metals & Machinery	High-Tech	Textiles	Food & Beverage	Other Services	Other Manufact.
$\ln(TRADE_{ijkt})$	0.250***	0.593***	0.262***	0.873***	0.347***	0.141***	0.510***	0.584***
	(0.0159)	(0.0311)	(0.0294)	(0.0226)	(0.0555)	(0.0203)	(0.0259)	(0.0247)
$\ln(FDI_{ijt})$	-0.0119	-0.0379	-0.0167	-0.0506*	-0.0416	-0.0156	-0.0369	-0.0155
	(0.0105)	(0.0223)	(0.0192)	(0.0212)	(0.0462)	(0.0281)	(0.0305)	(0.0169)
$MARKET_{ijt}$	-0.00351***	7.6E-06	-0.00553***	-0.000918	-0.00692	-0.00201***	-0.00533**	-0.00345***
	(0.000588)	(0.000888)	(0.00165)	(0.000899)	(0.00457)	(0.000583)	(0.00187)	(0.000861)
$WEALTH_{ijt}$	0.0512	0.0693	0.0190	-0.0634	0.234	-0.134	0.0310	0.144
	(0.0732)	(0.121)	(0.152)	(0.184)	(0.280)	(0.110)	(0.209)	(0.128)
$\ln(FVA_{jkt})$	0.126***	0.104	0.332***	0.129*	0.616***	0.0572*	0.421***	-0.0907**
	(0.0247)	(0.0605)	(0.0642)	(0.0582)	(0.174)	(0.0288)	(0.0358)	(0.0334)
$\ln(OVA_{ikt})$	0.0865***	0.585***	0.130**	-0.195*	-0.463***	0.287***	0.0698	0.121**
	(0.0221)	(0.0735)	(0.0486)	(0.0760)	(0.124)	(0.0439)	(0.0360)	(0.0432)
$\ln(IN\_FDI_{ikt})$	0.000859	-0.00976*	0.00387	-0.000748	0.160	0.117***	-0.00162	0.00372
	(0.00131)	(0.00464)	(0.00570)	(0.00362)	(0.139)	(0.0256)	(0.00188)	(0.00462)
$\ln(OUT\_FDI_{jkt})$	-0.00129**	-0.0136	0.00617	-0.000725	0.292**	0.0158	-0.00117	0.00849
	(0.000488)	(0.00710)	(0.0108)	(0.000838)	(0.0949)	(0.0162)	(0.00102)	(0.00799)
IPR_{jt}	19.47***	14.52**	21.64***	18.37*	20.36*	18.83***	19.71***	20.52***
	(2.363)	(5.503)	(3.797)	(7.853)	(8.434)	(4.109)	(4.711)	(3.064)
log Distance	0.0302	0.408***	0.000358	0.269***	-0.143	-0.223***	0.145**	0.186***
	(0.0209)	(0.0603)	(0.0339)	(0.0588)	(0.138)	(0.0493)	(0.0547)	(0.0317)
Border Dummy	0.409***	0.413*	0.213	-0.0205	0.103	0.171	0.118	0.201*
	(0.0785)	(0.180)	(0.115)	(0.180)	(0.249)	(0.116)	(0.148)	(0.0931)
Language	0.353***	0.215	0.521***	0.298*	-0.0119	0.371***	0.550***	0.219**
Dummy	(0.0563)	(0.117)	(0.0858)	(0.139)	(0.250)	(0.0968)	(0.102)	(0.0750)
Constant	-99.37***	-93.15***	-113.5***	-95.29*	-102.2*	-93.77***	-111.1***	-104.3***
	(11.62)	(26.66)	(18.49)	(38.34)	(41.07)	(19.92)	(22.88)	(14.92)
Log-likelihood								
Observations	59493	7344	7475	4414	3289	8257	14751	13963
Pseudo R^2	0.957	0.937	0.968	0.978	0.857	0.892	0.937	0.952

Robust Standard errors in parentheses, * p<0.05, ** p<0.01, *** p<0.001. All estimations use year-fixed effects.

Table 5: IV Regression of Bilateral Trademark and Patent Flows for OECD Countries, 2004 – 2008. Dependent Variable is Bilateral Trademark Flows from country *i* to country *j*.

	First Stage:		Second Stage:	
	(1)	(2)	(3)	(4)
	$\ln(TRADE_{ijkt})$	$\ln(FDI_{ijt})$	Trademarks – All NICE	Patents – All NICE
$\ln(OGDP_{it})$	0.679***	2.163***		
	(0.0788)	(0.0871)		
$\ln(FGDP_{jt})$	0.884***	0.600***		
	(0.0686)	(0.0759)		
$\ln(TRADE_{ijkt})$			0.351***	0.931***
			(0.00570)	(0.0139)
$\ln(FDI_{ijt})$			-0.517***	-0.539***
			(0.0603)	(0.106)
$MARKET_{ijt}$			-0.00355***	-0.00269***
			(0.000476)	(0.000411)
$WEALTH_{ijt}$			-0.258***	-0.182**
			(0.0150)	(0.0617)
$\ln(FVA_{jkt})$	0.0159*		0.117***	0.101***
	(0.00705)		(0.00900)	(0.0170)
$\ln(OVA_{ikt})$	-0.0297***		0.0166	-0.00593
	(0.00727)		(0.00926)	(0.0154)
$\ln(IN\_FDI_{ikt})$			0.00242***	0.00181
			(0.000615)	(0.00130)
$\ln(OUT\_FDI_{jkt})$			0.00261***	-0.000887
			(0.000268)	(0.000475)
IPR_{jt}			-1.614***	19.93***
			(0.165)	(2.293)
log Distance	-1.370***	-1.151***	-0.688***	0.574***
	(0.00760)	(0.00838)	(0.0705)	(0.125)
Border Dummy	0.325***	0.373***	0.345***	0.589***
	(0.0156)	(0.0214)	(0.0303)	(0.0870)
Language Dummy	0.259***	0.506***	0.669***	0.365***
	(0.0170)	(0.0206)	(0.0420)	(0.0744)
Constant	-19.49***	-62.51***	17.16***	-105.6***
	(3.204)	(3.556)	(1.390)	(11.23)
F-statistic	12394.46	517.98		
R^2	0.769	0.748		
Log-likelihood				
Pseudo R^2			0.746	0.941
Observations	59493	59493	59493	59493

Robust Standard errors in parentheses, * p<0.05, ** p<0.01, *** p<0.001. All estimations use year-fixed effects, along with origin and destination country fixed effects. First-stage is estimated using a panel fixed-effects model. Estimation (1) uses NICE class fixed effects.

Appendix

I. Additional Details Regarding the ALP Matching Methodology

A. Comparison of USPTO-Based and ROMARIN-Based Class-Level ALP Concordances

The process we devised for constructing the 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 by 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). We constructed Class-Level concordances for both SITC and ISIC using both USPTO and ROMARIN data and can compare these concordances to better understand potential differences. In this subsection, we directly compare the resulting ALP concordances for SITC. Later, we indirectly compare them based on differences in TM use intensity in the case of Vietnam.

Overall, the SITC concordance based on USPTO data is remarkably similar to the one based on ROMARIN data. In a given NICE Class, the top ranked SITC 2-digit matches are nearly always identical. Lower ranked matches tend to differ slightly, but these are mostly inconsequential due to their relatively low weights. To test our hypothesis that a USPTO-based concordance is less noisy than a ROMARIN-based concordance, we can compare the profile of estimated weights; a relatively noisy concordance will yield more matches to a given NICE Class with lower weights. When we do this for our SITC concordances, we find some weak evidence that the USPTO-based concordance is more precise. On average, 7.8 SITCs (2-digit) match to each NICE Class (after applying the 2% cutoff) based on USPTO data, whereas 8.6 SITCs match to each class when ROMARIN data are used.

We can test this hypothesis more rigorously by computing the average weight across NICE Classes for the *n*th ranked SITC. The average weight for the 1st SITC match is 38.6% when based on USPTO data and 37.8% when based on ROMARIN data. Although this same pattern – relatively lower average weights for ROMARIN-based than for USPTO-based concordances – persists across other rankings (i.e., 2nd, 3rd, etc.), these weighting profile are statistically indistinguishable. The surprising similarities between the USPTO-based and ROMARIN-based ALP concordances suggest that differences in defensive registration of goods and services are largely washed away in the processing and aggregation of our matching algorithm.

II. Pre-defined Country Groups

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

Table A.1: List of Countries by Income Group (by ISO Country-Code). Source: World Bank Classification

Low TM Intensity	Medium TM Intensity	High TM Intensity
JPN, KOR, HUN, CHL, MEX, SWE, CZE, ISR, GRC, DNK	CAN, NOR, FRA, USA, ISL, AUT, ITA, POL, SVN, FIN, GBR	PRT, ESP, DEU, AUS, CHE, NLD, BEL, SVK, EST, NZL, IRL, LUX

Table A.2: List of OECD Countries by TM Intensity. Low Intensity countries are those with domestic TM intensities less than 20x OECD weighted average. Medium TM Intensity countries are those with domestic TM intensities between 30 and 100x OECD weighted average. High TM intensity countries are those with 100x ore more OECD weighted average. Note that OECD weighted average was weighted using a country's value-added and is relatively low due to Japan and Korea's low TM intensity and large value-added.