The Global Footprint of Local Corporations

Stephen Petrie, Trevor Kollmann, Russell Thomson, Alexandru Codoreanu, Elizabeth Webster
Swinburne University of Technology

Abstract

The measurement of firms pursing new goods producing business models, often called fabless manufacturing or own-brand importer-marketers, has centered on industry-specific micro measures that are limited by cost, confidentiality, coverage or geographic granularity. Most often these firms that adopt these factoryless product-focused models are unhelpfully classified as wholesale traders. We introduce a new internationally linked trademark database (*TM-Link*) and show that variation in international trademarking activity, after controlling for exports and outward FDI, conforms well to product complexity measures and existing evidence on the growth of both fabless manufacturing and production fragmentation (by State and by industry). We argue that since companies trademark in markets where they want control over product management, design and distribution, domestic and international trademark data can shed light on the firms and locations that are engaged in these new business models.

INTRODUCTION

It is emblematic of modern production that managers in Cupertino, California have de facto control over the daily activities of more than one million employees of Taiwanese owned Foxconn in Shenzen, China. However, little of Apple's vast global footprint is readily attributable to its home region of Silicon Valley using traditional trade and investment indicators (see e.g., Xing 2020).[1] Brand manufacturers, such as Apple, lead in the design, development and production of goods which typically carry the companies' branding. Fabrication is increasingly undertaken offshore via arms-length contracts. In many cases, firms' global footprints cannot be readily observed from exports or foreign direct investment (FDI). Since firms are classified according to the activity which contributes the largest share of their value-added, firms conforming to these emergent business models are typically classified within the wholesale trade sector by national statistics agencies. New measures are needed to better understand the spatial distribution of design, management, coordination and control which is now geographically and organizationally separated from fabrication and assembly. In this article, we present a new dataset of US firms' international trademark filings (called TM-Link) and demonstrate that these

[1] For example, if the value added of Apple IP were to be recorded as a US export, it would contribute 3.7 per cent to total United States exports (Xing 2020).

data provide a novel perspective on the global footprint of US companies.[2] We show that international trademark filings provide a new and readily available tool to identify fabless goods producers classified in the wholesale industry.

These emerging business models, including Global production sharing[3] and fabless or ‘factoryless’ goods producing firms, are central to understanding and measuring regional patterns of economic development (Jones and Kierkowski 2000; Gereffi et al. 2005; Baldwin and Venables 2013; Bernard and Fort 2015; Morikawa 2016). In many of these new business models, fabrication of products is outsourced, often offshore, but the firm maintains control of the production process, owns the associated intellectual property and bears the entrepreneurial risk. Critical to their business model is the capture of profits through their control over the whole course of production: knowledge, ownership of materials, design and quality. Important loci of control are ownership of brand and distribution channels (Knight and Liesch 2016; Hertz and Thomson 2021). Regions benefit from high-value added managerial and coordination activities. Additionally, agglomeration of production management provides a conduit to high-return investments and pathways to market for local innovators – regardless of the location of production.[4]

To date, measurement of these new business models has centered on industry-specific micro measures of offshored outsourcing, dedicated surveys; (confidential) tax records; service exports; and IP sales and royalties (Bernard and Fort 2015; Bayard, Byrne and Smith 2015; Johnson 2018). Although these data are limited by confidentiality, coverage or geographic granularity, the collated evidence indicates that fabless international business models are widespread and growing. Bernard and Fort (2015) estimate that by 2007, 12 per cent of firms in the US wholesale industry were factoryless goods producers. The US Census Bureau estimates that manufacturing value-added would have been 5 to 20 per cent greater in 2007 if all factoryless goods producers were reclassified to manufacturing (Bayard, Byrne and Smith 2015). Tomiura (2018) cites several national

[2] TM-Link is publicly and freely available online at https://www.tmlink.net.au/

[3] Global production sharing (which is also referred to as global production fragmentation) refers to 'slicing' up of production processes into a series of geographically separated stages to take advantage of differing factor cost advantages.

[4] For instance, a well-known barrier to aspiring innovators operating on the periphery as prospective innovators must reach key decision in these regions.

surveys of international outsourcing but notes their limitations with respect to sample bias.[5] Morikawa (2016) estimated that 19 per cent of Japanese product producing firms were fabless, of which one in six were internationally active. Most were nominally classified in ICT services or wholesale trade though they invested intensively in intangible assets such as R&D. The emergent fabless, or intangible, business model includes firms across a wide range of sectors, from toys and automotive to biotech, footwear and even cookie production (see Bernard and Fort 2015; Morikawa 2016).[6] Many are classified by national statistics agencies as being in the wholesale sector, since this reflects their main observable activity.

Exports, foreign direct investment (FDI) and intellectual property royalties are often used to measure the internationalisation of production. Each tell part of the story, but each have limitations. Export data for instance do not capture goods assembled in third countries. Additionally, gross export data overestimate a country's international activity because partly finished goods are shipped back and forth across borders, often with little domestic value-added. Although some thought has been given to using census and customs data to build Input-Output tables for exporting versus non-exporting firms (e.g. Johnson 2018), value-added trade data will not illuminate the ultimate controllers of global value chains, nor will it capture the intangible business model exemplified by Apple Inc. International royalty payments and trade in intellectual property also tell part of the story, though these are not widely available at the industry or firm level. Even where they are available, royalty payments are subject to a range of limiting caveats. It is perennially hard to verify whether the price is a true reflection of the value of intangible assets, especially where trade occurs between affiliates. Evidence suggests 5 per cent of the revenue from smartphone sales is tagged as a return to IP but a further 40 per cent is captured via the high retail price margin (WIPO 2017).[7]

There may never be an 'ideal' indicator of the reach of fabless US companies, but we argue that domestic and international trademark data tell an important part of the story and have a range of compelling advantages.

[5] There are datasets in France, the US, Italy and Ireland.

[6] E.g. Nike, Reebok, The Limited, Gap (see Xing 2017), Solectron (see Gereffi, Humphrey and Sturgeon 2005). For the discussion of dry labs in biotechnology (see Nicol, Liddicoat and Critchley 2013). Carman's Fine Foods, Saltwater Seafoods. See also Herz and Thomson (2021)

[7] In contrast, Neubig and Wunsch-Vincent (2017) estimate that tax-induced mis-measurement may overstate the true flow of intangible inputs by 35 per cent – even more for high tax countries.

Trademarks are distinctive signs which identify which goods or services are provided by specific people or companies (WIPO, 2004). There is a strong commercial imperative for companies desiring to launch a new product to obtain their own trademark for that market (De Faria and Sofka 2010; Giarratana and Torrisi 2010; Sandner and Block 2011; Nasirov 2018). Trademarks are relatively cheap to obtain[8] and provided their text and symbols are distinctive relative to existing marks in the target market, applications are almost always granted (Graham et al. 2013). An exclusive trademark underpins incentives for companies to make ongoing investments into product quality, content and distribution networks (Economides 1988; Lancaster 1990; Mendonça et al., 2004; Jensen and Webster 2004, 2009; Fosfuri and Giarratana, 2009; Krasnikov et al., 2009; Flikkema et al., 2014; Flikkema et al., 2019; Castaldi 2020).

To the best of our knowledge, domestic and international trademark filing data has not been used to identify global business activity, despite their potential advantages. Trademarks are publicly-available, firm-level micro data and are therefore highly geographically and industry specific. They give the identity of the business and are also timely. This stands in contrast to publicly available wholesale trade, export and FDI data which tend to be geographically and or industrially aggregated; published with time delays; de-identified or when available as micro data, and are only available to trusted local analysts with substantial legal restrictions. Trademarks also cover a wide range of economic activity. Although many trademarks are associated with consumer products, they are also applied to intermediate inputs and capital goods. For example, Malmberg (2005) shows that between 1945 and 1960, Ericsson trademarked half of their professional[9] and component[10] products as well as every consumer product.[11]

[8] The United States Patent and Trademark Office charges $275 to obtain trademark registration, only a few hundred dollars after five years and another few hundred dollars every ten years.

[9] This included telephone equipment with a more professional profile (loudspeaker phones, answering/announcing machine), private branch exchanges, intercom phones, telex equipment, mobile radio and other electrical products (electric meters, electric fencing etc.)

[10] This comprised capacitors and vacuum tubes (including early transistors). Malmberg (2005) also finds AGA also trademarked intermediate goods.

[11] They did not trademark infrastructure such as goods for telecoms, railway operators and the military because customers were fewer and the sales process was more personal.

We have developed a comprehensive database of international trademark applications, filed by US firms in five foreign jurisdictions that are ‘equivalent’ to their domestic marks (called TM-Link). Trademark equivalents are trademark applications filed in different legal jurisdictions by the same company to protect the same product or service. To identify trademark equivalents, we developed a neural-network linking algorithm. Our approach reveals that about 12 per cent of US trademarks filed by US domiciled firms have an equivalent in one of the other five jurisdictions.

We assembled a State-by-industry dataset including international trademark share, exports and outward FDI to show the role international trademarking data can play in illuminating international business activity of fabless product firms in the USA. The data confirm that the share of international trademarking by firms domiciled in the USA is strongly associated with exports and outward FDI. After controlling for exports and outward FDI, the remaining variation in international trademarking conforms well to existing evidence on the growth of both fabless manufacturing and production fragmentation (by State and by industry). We then take a more detailed look at firms classified as wholesaling firms. An analysis of available corporate documentation indicates that wholesale firms with substantial international trademark portfolios are lead firms in global production networks. Finally, we consider auto manufacturers which is an illustrative case study of an industry that has extensively served foreign markets through globalized and arms-length manufacture.

DATA

To test the relationship between international trademarking and the global footprint of corporations, we have constructed a novel dataset incorporating trademarks, export data, and outward FDI data. Data from three main sources were linked.

- Linked equivalent international trademark applications by US domiciled firms for the period 1996-2018.
- Exports by industry-State-level (4-digit NAICS), from US International Trade Administration database (2003-2018).

- Outward FDI series at the industry-state-level (4-digit NAICS), from the Financial Times fDi Markets (2003-2018).

To identify international trademarks owned by domestic firms, we developed a neural-network linking algorithm which identifies similar trademarks across different countries. We link trademark applications from the USA, Canada, Australia, New Zealand, the United Kingdom and the European Union. Trademark equivalence is assessed through algorithmic comparison of the following: trademark text (the words or text depicted in the trademark);[12] filing date; Nice classification; and applicant name. Once we have cleaned the data from each national office and consolidated duplicate records,[13] we algorithmically identify equivalent trademarks across jurisdictions – i.e. families of trademarks from a given firm filed across different national offices. The linking algorithm we developed to identify these families was applied to all applications that contain trademark text.[14] As described in Appendix A, our linking algorithm involves two main facets: a binning algorithm and neural network classification algorithm (further details of the linking algorithms can be found in Petrie & Julius 2019; and Petrie et al. 2020).

To assess the accuracy of the linking algorithm, we manually examined a random sample of asserted positive links and found that 98 per cent of those links were true positives and 2 per cent were false positives. We successfully obtain approximately 85 per cent of a priori known links (i.e. TM-TM pairs that share the same international registration number). Table 1 presents data on USPTO trademark applications filed by US domiciled firms, from 2003 to 2018. Of the 4.7 million trademark applications, 64,513 have an international equivalent identifiable via the Madrid Protocol. The figure is only marginally improved by looking at trademark applications with international legal priority, which identifies 69,261 international marks (equivalent

---

[12] Trademarks include word and text as well as specific design or typographic elements. We do not consider the associated trademark images (stylistic elements and visual designs).

[13] Multiple applications for the same identical trademark can reflect recording errors or, in some cases, multi-Nice Class trademarks would require separate applications each with a single Nice Class. Recording error duplicates, where all fields contain identical data, were removed. For some offices, older applications could only be filed under a single Nice class. Thus, any application covering multiple Nice classes would be submitted as several duplicate applications with different application ID numbers, each with a different Nice class. To rectify this, we create a string that identifies such duplicates by concatenating the filing date, word mark text, and applicant (with any non-alphabet characters removed from applicant text and all letters converted to upper case) into a single string. We then remove an application if its duplicate string is identical to that of any other application.

[14] Ninety-six percent of all trademark applications include text.

applications identifiable via the Madrid Protocol are a subset of those with international legal priority). By contrast, our neural network algorithm identifies 563,961 US trademarks with international equivalents. That is, our approach reveals that about 12 per cent of US trademarks filed by US domiciled firms have an equivalent in one of the other five jurisdictions.

Table 1: Summary of internationally-linked trademark applications

Trademark applications	US applications by US domicile firms, 2003 to 2018
Total	4,695,583
with international equivalents indicated by Madrid	64,513
with international equivalents indicated by Legal Priority	69,261
with international equivalents indicated by TM-Link	563,961

Source. TM-Link. The TM-Link data is available online at https://www.tmlink.net.au/

Even though TM-link are available at the micro-level, to evaluate the relationship between international trademarks and the global footprint of US firms the data is aggregated to 4-digit NAICS industry group across the 50 States from 2003 through 2018. To map between Nice Classification (Trademarks) and NAICS Classification (FDI and Export data), we use a concordance table based on full matching between USPTO trademark applications and Bureau van Dijk's Orbis database.[15] Trademark applications, by industry group, were then aggregated to the State level and merged with the State-industry level data on exports and outgoing FDI.

Outward FDI are compiled from data provided by fDi Markets. Compiled by the *Financial Times*, these data reflect a comprehensive census of all FDI projects globally from 2003 to 2018 and cover projects with a total value of approximately $10 trillion USD over the period.[16] fDi Markets data are collected via media, direct market research, and survey of more than 2,000 industry organizations. Each investment project included in the dataset is cross-referenced and validated using multiple sources.

[15] Based on 4,114,075 USPTO trademark applications filed between 1996 and 2013 linked to the corresponding record among the 25.6 million U.S firms listed in Orbis. The matching process, described in Appendix B.

[16] Aggregate flows reported in fDi Markets data are very similar to the total flows reported in OECD

To measure exports, we use “State of Origin of Movement” export series, 2003-2018, are reported at the industry-State level (4-digit NAICS), Exports are coded to the State of origin not the port of export.[17] Data is restricted to firms in the manufacturing sectors (31-33), however as discussed below, we assign a value of $1 to exports within the wholesale sectors (42) for the purposes of the statistical analysis.

INTERNATIONAL TRADEMARKS AND THE GLOBAL FOOTPRINT OF US FIRMS

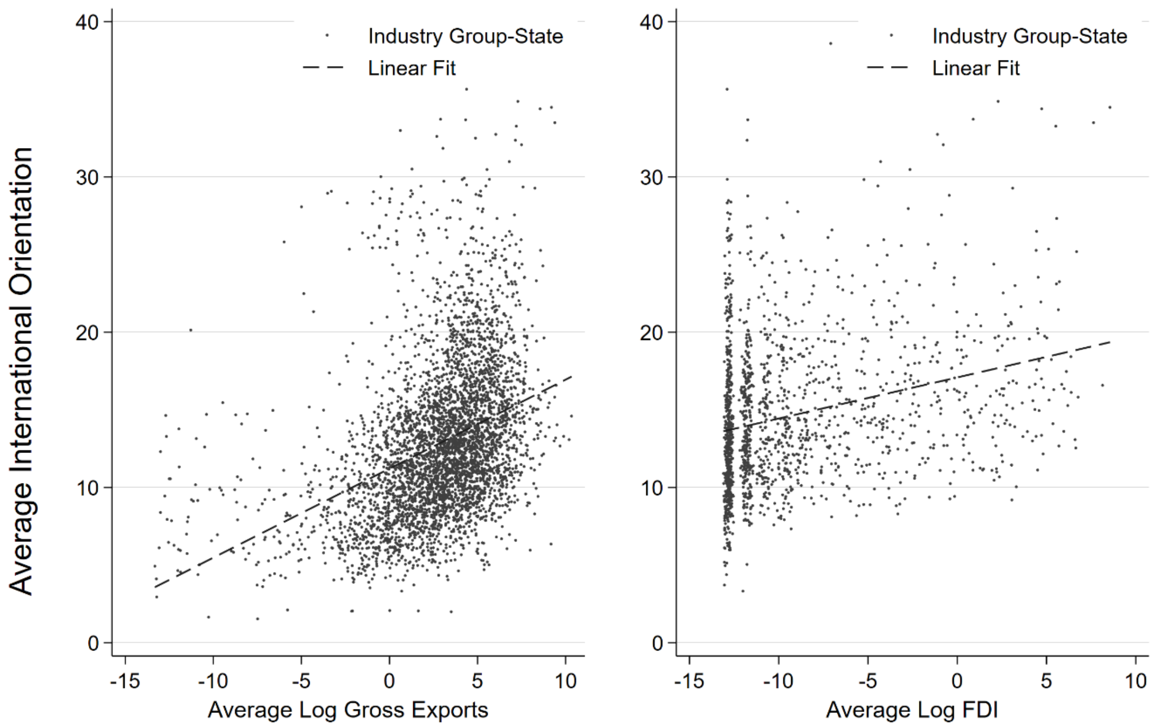
Our primary interest in modelling the association between international trademarks is to first establish the extent to which it reflects traditional trade and investment measures of globalization (export and FDI activity); and second to demonstrate the extent to which the additional information provided by international trademarks sheds new light on unmeasured level of global economic activity, that is, innovations in ‘fabless’ goods producing firms which are often classified as wholesaling by national statistics agencies.

As previously discussed, a trademark with an international-link signals a decision for a firm to expand into the foreign market. If the firm is a traditional integrated manufacturing firm, the internationally-linked trademark would signal that the firm (or domestic contractor) is either exporting its production or is investing in capital projects overseas.

Figure 1 shows two scatter plots of the relationship between the average share of trademark applications with an international link (international trademark intensity) in a 4-digit NAICS industry-group and State between 2003 and 2018 and its corresponding logged exports and FDI. The positive correlation between the international trademark intensity to log exports and log FDI confirms that firms expanding their international footprint via these traditional pathways tend to seek to protect their branded products in foreign markets. Yet the low level of collinearity (correlations are below 0.3) suggest that these are not sufficient at explaining international trademark intensity for US manufacturing firms – there is more to international expansion than is explained by these traditional measures alone.

[17] The origin of movement may not be the origin of production if the goods are consolidated along the journey. This is apparent in agriculture data that overestimates the impact of agriculture output for Louisiana and other agricultural ports. This may lead to a reduction in real exports relative to trademarking activity in agriculture states. In practice the inclusion / exclusion of agricultural sectors from our analysis does not affect our main findings.

Figure 1: Scatterplot relationship comparing International Trademark Intensity to Exports and FDI



Note. Average International Orientation measures the average share of trademarks between 2003 and 2018 in an industry group-state that have an international link within the TM Link database. Gross Exports and FDI are adjusted for inflation using the Consumer Price Index for all urban consumers (CPI-U) with 2012 set as the base year. Both measures are averaged between 2003 and 2018. The unit of observation is industry group (NAICS, 4-digit) and state.

We want to explore whether offshore manufacturing may bridge that gap seen within Figure 1. We can do this by modelling the ratio of international trademarks to total trademark applications (International Orientation) within a State-industry group as a function of exports, FDI and off-shore manufacturing:

$$\frac{International\ TMs}{TMs} = f(Exports, FDI, Offshore\ 3rd\ Party\ MFG)$$

If there are increasing marginal returns to each international trademark, such that at higher rates of export or FDI, fewer additional international marks are required, then a semi-log format is more appropriate. As we do not observe offshore 3rd party manufacturing, the extent will be captured in the error term, viz:

$$\left(\frac{International\ TMs}{TMs}\right)_{ist} = \alpha + \beta_1 \ln(Exports)_{ist} + \beta_2 \ln(\mathrm{FDI})_{ist} + \varepsilon_{ist} \quad (1)$$

$$\varepsilon_{ist} = \beta_3 \ln(Offshore\ 3rd\ Party\ MFG)_{ist} + \omega_{ist}$$

where i is the four-digit NAICS industry group, s is the US State and t is year. We additionally add a quadratic term to both log exports and log FDI to control for the non-linear relationships between them and international trademark intensity.

As we are particularly interested in investigating what these new international linked trademark data reveal about factoryless goods producing firms that might be 'misclassified' in the wholesale sector, we augment the model to include State-level establishments in the wholesaling industry. In addition, we control for the number of establishments within the industry and State to control for changes in scale of the industry which may influence trademarking behavior. As robustness checks for the relationships, we also include year, State, and industry fixed effects. We include all industries within the manufacturing and wholesaling sectors (NAICS 31-33, 42). As wholesaling has no associated exports in the census data, we assign a value of $1 to the exports.[18]

As offshored, 3rd-party manufacturing is not observed, a large, positive residual will signal those State-industry groupd that are active in new international business models.

Table 2 presents a summary of the data used in the regressions (noting that the unit of analysis is 4-digit industry group and State from 2003 through 2018). The average number of trademark applications for an industry group within a year is 19.6 of which 2.6 are international marks. The international share of trademarks was 12.5 per cent, slightly lower than the ratio of the averages. Average gross exports are $219.1m (2012 prices) and average outgoing FDI flows were $13.1m (2012 prices). The average number of establishments in an industry group within a state was 148.8, while the average number of wholesale establishments within a State was 12,253.

[18] The inclusion of the wholesaling industry groups within the regressions does not impact the rankings of the manufacturing industry groups discussed below.

Table 2: Summary Statistics of the Regression Sample

	Mean	Std Dev	Min	Max
International Trademark Applications	2.6	8.2	0.0	280.8
Trademark Applications	19.6	63.0	0.0	2,087.6
Share of TM Applicns with Intl Links	12.5	6.8	0.1	93.6
Real Gross Exports	219.1	1,136.1	0.0	59,616.3
Real Outward FDI	13.1	204.1	0.0	21,583.1
Wholesale Establishments	12,253.1	11,978.3	663.0	64,391.0
Establishments	148.8	399.4	1.0	7,934.0

Notes. Gross Exports and Outward FDI are adjusted for inflation using the Consumer Price Index for all urban consumers (CPI-U) with 2012 set as the base year. See Appendix C for precise variable definitions and sources. N = 81,600.

Table 3 presents the estimations of Equation (1). As shown in Column (1), export and FDI activity is strongly positively correlated with the international trademarking intensity. In addition, the relationship suggests that relationships increase at an increasing rate. This result holds when we include control variables for both within industry-group and wholesale establishments as well as a series of State, year and industry fixed-effects [Columns, (2) – (5)]. Overall, these findings are consistent with agglomeration which would suggest that firms within a region and industry that was already globally-focused are more able to leverage that capacity.

Table 3: Dependent variable = log (share of international trademarks), USA, 2003-2018, OLS

	(1)	(2)	(3)	(4)	(5)
Log Real Gross Exports	0.390***	0.421***	0.287***	0.292***	0.145***
	(0.007)	(0.006)	(0.006)	(0.006)	(0.010)
Log Real Gross Exports Squared	0.028***	0.031***	0.019***	0.020***	0.009***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Log Real Outward FDI	0.271***	0.203***	0.160***	0.154***	0.204***
	(0.025)	(0.023)	(0.019)	(0.018)	(0.017)
Log Real Outward FDI Squared	0.016***	0.011***	0.011***	0.010***	0.015***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Log Wholesale Establishments				4.569***	4.438***
				(0.391)	(0.354)
Log Establishments				-0.009	0.552***
				(0.015)	(0.030)
Constant	11.685***	14.882***	11.867***	-29.570***	-30.668***
	(0.147)	(0.155)	(0.172)	(3.551)	(3.224)
Year Effects	N	Y	Y	Y	Y
State Fixed Effects	N	N	Y	Y	Y
Industry Fixed Effects	N	N	N	N	Y
Adj. R-squared	0.066	0.243	0.509	0.516	0.603
N	81,600	81,600	81,600	77,742	77,742

Note. * p < 0.10, ** p < 0.05, *** p < 0.01. Dependent Variable: Share of Trademark Applications with International Links (International Orientation of Industry). Unit of analysis is Industry Group (4-digit NAICS)-State-Year. Sample are industry groups within the manufacturing subsector (NAICS 31-33) and wholesaling subsector (NAICS 42) between 2003 and 2018 (inclusive). Log transformations add one to the value to avoid dropping observations with no exports, FDI, etc. Gross Exports and FDI are adjusted for inflation using the Consumer Price Index for all urban consumers (CPI-U) with 2012 set as the base year.

We expect that patterns in the size of the average residual from Column (1) estimations to point to those regions and industry groups which are both engaged in high levels of offshore, arms-length manufacturing. Table 4 shows the average residual value for all industry groups across the nine US Census Divisions. The East North Central Division, which includes the rust belt States of Michigan, Ohio and Indiana have the largest residual indicating that many of firms domiciled in that division are likely engaging in offshore manufacturing. New England has the second highest average residual. The lowest residuals are in the West South Central and Mountain regions (see Figure 2). Overall, Table 4 and Figure 2 highlight that whereas regional differences exist, States within each region can also vary. In particular, the results suggests that the States that have had long histories of manufacturing and stronger union laws are also the States that are shifting to new business models for manufacturing characterized by offshoring and arms-length international outsourcing.

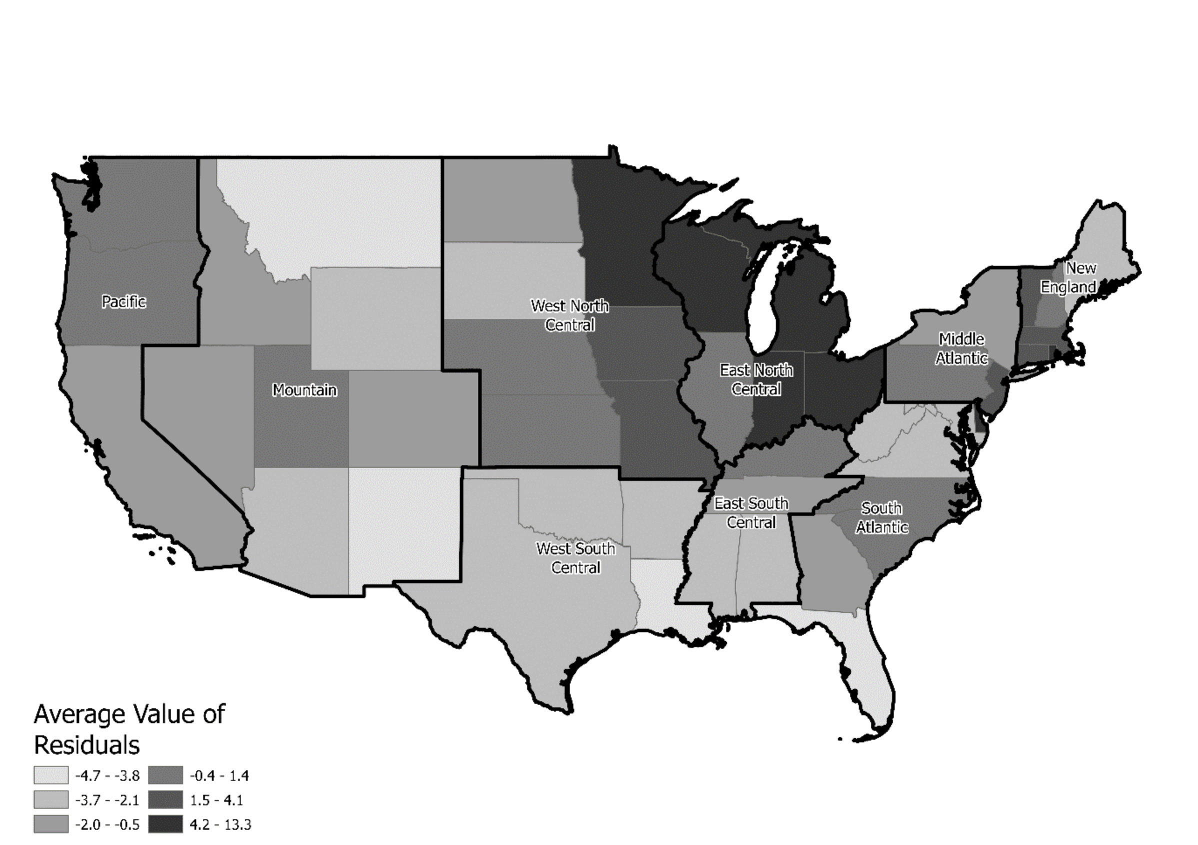
The regional differences amongst the residuals closely mirror Fort, Pierce and Schott (2018) who found that manufacturing employment since the late 1970s fell the most amongst the firms located in New England, Mid-Atlantic and the East North Central divisions. Crucially for our purpose, their results also found that manufacturing firms continued to increase their employment amongst non-manufacturing plants. This supports the view that internationally linked trademarks provide a window into agglomeration of non-manufacturing activities of goods producing firms.

Table 4: Average Residual Value by U.S. Census Division and Selected Summary Statistics (Mean Value)

Census Division	Residual	International Trademarking Intensity	Exports	Outgoing FDI
New England	2.01	14.08	79.82	7.62
Mid-Atlantic	0.65	14.38	383.79	33.21
East North Central	4.46	17.98	369.58	37.79
West North Central	0.97	13.14	78.28	4.54
South Atlantic	0.21	12.97	187.56	7.05
East South Central	-1.72	10.93	169.99	2.31
West South Central	-3.13	9.61	573.98	17.07
Mountain	-1.97	9.81	61.46	3.01
Pacific	-1.92	10.53	391.71	25.95

Note. The Residual column indicates the average value of the residuals across industries within a U.S. Census division derived from Column (1) in the primary regression results in Table 3. See Appendix C for variable definitions. Rankings are robust to the specification.

Figure 2: Average Residual Value by U.S. State



Note. Residual indicates the average value of the residuals across industries within a U.S. State derived from Column (1) in the primary regression results in Table 3. Dark lines represent U.S. Census Division boundaries.

Table 5 provides the average residual by selected 4-digit industry groups. It shows that Other Transportation Equipment Manufacturing; Motor Vehicle Body and Trailer Manufacturing; Electric Lighting Equipment Manufacturing; and Ventilation, Heating, Air-Conditioning, and Commercial Refrigeration Equipment Manufacturing had the highest residuals and therefore are predicted to be most heavily engaged in offshore, 3rd party manufacturing global business models. Also as expected, the results indicate that the least likely to engage in these business models are commodity producers such as: Animal Slaughtering and Processing; Printing and Related Support Activities; Pulp, Paper, and Paperboard Mills and Fruit and Grain and Oilseed Milling The full set of average residuals is included in Appendix D.

It is also noteworthy that the rankings of residuals by industry group are related to product complexity. Felipe et al (2012) finds that product complexity, as measured by Hidalgo and Hausman's method of reflections, finds that chemicals, machinery and transportation goods amongst the most complex products, whereas textiles, foodstuffs and leather amongst the least complex. Although the correlation is not perfect, it nonetheless suggests that industries which manufacture complex goods are amongst those that find partnerships with overseas 3rd party manufacturing firms.

Although our rankings align with product complexity, they contrast with sector rankings of fabless manufacturing S&P 500 firms developed from a survey undertaken by Bayard, Byrne, and Smith (2015). The latter find that most firms selling toys and games, apparel, electronic components, computers and communications equipment, pharmaceuticals, and food companies were engaging with factoryless manufacturing in 2012. We note that the Bayard *et al*. survey industries are outsourcing fabrication and assembly to either or both domestic and foreign businesses. Insofar as they are reliant on domestic contractors for production, their subsequent exports would be captured within our data and would align the industries amongst those closer to traditional international business models which rely on directly exporting produced goods.

This argument is indirectly supported in Doherty (2015) which examines the impact on statistical measures for manufacturing classifications in NAICS if factoryless wholesale firms which import their own products would be integrated into the manufacturing classifications. They find that Food and Beverages amongst the subsectors least likely to be impacted, suggesting that food manufacturing in the United States is not reliant on offshore, 3rd party manufacturers. However, the same argument cannot be used for apparel as retail brands are increasingly using FDI to build an international retail presence, thus it is possible that this FDI is leading to smaller residuals than would be expected.

Table 5: Average Residual Value for Selected Industry Groups (4-Digit NAICS)

Rank	No	Industry Group (NAICS)	Residual	International Trademarking Intensity	Exports	Out FDI
1	3369	Other Transportation Equipment Manufacturing	4.4	16.9	67.1	8.7
2	3362	Motor Vehicle Body and Trailer Manufacturing	4.2	16.9	143.8	1.8
3	3351	Electric Lighting Equipment Manufacturing	3.9	16.2	43.2	4.3
4	3334	Ventilation, Heating, Air-Conditioning, and Commercial Refrigeration Equipment Manufacturing	3.4	16.6	144.2	2.7
5	4231	Motor Vehicle and Motor Vehicle Parts and Supplies Merchant Wholesalers	2.9	14.0	0.0	1.1
6	3365	Railroad Rolling Stock Manufacturing	2.9	14.9	52.6	3.0
7	3366	Ship and Boat Building	2.8	15.0	51.9	1.5
8	3352	Household Appliance Manufacturing	2.7	15.7	129.3	4.6
9	3335	Metalworking Machinery Manufacturing	2.7	15.7	140.0	1.9
10	3363	Motor Vehicle Parts Manufacturing	2.5	16.8	968.3	67.1
93	3113	Sugar and Confectionery Product Manufacturing	-2.6	9.7	45.8	10.5
94	3115	Dairy Product Manufacturing	-2.6	9.8	83.6	4.2
95	4251	Wholesale Electronic Markets and Agents and Brokers	-2.8	8.2	0.0	0.0
96	3114	Fruit and Vegetable Preserving and Specialty Food Manufacturing	-3.2	9.3	100.9	6.2
97	3112	Grain and Oilseed Milling	-3.2	10.1	229.9	9.0
98	3119	Other Food Manufacturing	-3.4	9.8	145.1	10.1
99	3221	Pulp, Paper, and Paperboard Mills	-3.4	9.7	265.7	12.1
100	3118	Bakeries and Tortilla Manufacturing	-3.4	8.7	31.6	1.8
101	3231	Printing and Related Support Activities	-4.0	8.9	119.4	3.8
102	3116	Animal Slaughtering and Processing	-4.4	9.0	342.2	2.6

Note. The Residual column indicates the average value of the residuals within a 4-digit NAICS industry derived from Column (1) in the primary regression results in Table 4. See Appendix C for variable definitions. Rankings are robust to the specification. The complete list of residuals by industry can be found in Appendix D.

As we saw in Table 4, regional differences exist across all industry groups, thus we should expect those regional differences to persist within industry groups. To visualize these differences, Figure 3 plots the average residuals in an industry group against the corresponding average share of trademarks in that industry group for all observations within a U.S. Census Region and 3-digit NAICS subsector (Northeast, Midwest, South and West). The vertical line was chosen based on the constant value in the linear regression, but it is not intended to be a strict demarcation of industries within each classification.

The share of trademarks with international links roughly corresponds to the international orientation of that industry-region. Higher levels suggest more firms within that region are engaged in overseas business. As suggested by Model (1), a high residual corresponds to industry groups engaging in the new emerging global business models which utilize offshore, fabless manufacturing. Industries that have both a high residual and are highly integrated internationally as measured by the share of trademarks with international links, we would classify them as the leading industries in these emerging business models. Conversely, if they are highly integrated internationally, but have a small residual, that would be indicative that the industry is still reliant on traditional exporting, FDI and domestic contract manufacturing. Lastly, industries which are not as highly integrated internationally and have low residuals are those with a more domestic focus and those without typically strong branding such as commodities.

Figure 3: Relationship between International Orientation of Industries and Global Business Models by US Census Region

Note. Residual indicates the average value of the residuals within a 3-digit industry group-state within a U.S. Census Region and 3-digit NAICS subsector based on the results from Column (1) in the primary regression results in Table 3. The Northeast Census Region corresponds to the New England and Middle Atlantic Census Divisions, the Midwest Census Region corresponds to the West North Central and East North Central Census Divisions, the South Census Region corresponds to the West South Central, East South Central, and South Atlantic Census Divisions, and the West Census Region corresponds to the Pacific and Mountain Census Divisions.

Figure 3 further highlights that the regional differences found in Table 4 cannot be fully attributed to compositional differences in industries. Even amongst subsectors with strongly positive residuals such as Transportation Manufacturing (336), significant variation exists amongst the regions. Firms within Midwest transportation manufacturing appears to be more connected to modern international business models when compared to their Southern and Western counterparts. This is again consistent with Fort, Pierce and Schott (2018) which strongly suggested that the historic manufacturing bases in the Northeast and Midwest were amongst the firms to begin to diversify their businesses to include non-manufacturing jobs and shift away from in-house manufacturing.

Properties of key international trademarking firms

To obtain more granular evidence that internationally linked trademarks provide a novel window into the nature of firms' operations, we took a closer look at key firms. We undertook a detailed examination of the top 50 US firms which hold the largest portfolio of internationally linked trademarks but that are not traditional manufacturers. We first identify those US owned firms classified in Bureau van Dijk as being primarily in the wholesaling sector. We then use Bureau van Dijk global ultimate owner tables to exclude firms that have an ultimate owner firm with a primary classification of manufacturing. Finally, we search Google Patents to identify whether each firm in this list owns at least one patent, since registering patents is not consistent with the firms being mere wholesalers (i.e. brokers between agents).

These top 50 firms span a number of different industries, including apparel, consumer electronics, cosmetics, home appliances, pharmaceuticals, *inter alia*. Our prior is that these firms – all of which are classified as wholesalers – are product-focused firms with strong involvement in contract or offshore manufacturing. They are firms which lead in the design development of consumer products made using fabless business models and arms-length international contract manufacturing. We argue that international trademark ownership will provide a powerful first pass indicator of such firms.

For all firms in the top 50 we examined company websites and/or annual reports to identify whether the firm claims to lead product design and subsequent contract or offshore manufacture. Company statements were then verified using information from Dunn and Bradstreet's company database where possible. The top 10 firms are shown in Table 7 below. Of the top 50, we find that all but 3 firms are heavily engaged in product design, indicating these firms control product development and manufacturing – even in cases where they do not themselves own physical manufacturing plants or equipment assets. At least 70 per cent of firms in the list have arms-length contract manufacture. These findings support the view that international trademark applications provide a novel indicator that these firms control or lead global production networks.

Table 7: Top 10 US wholesaling firms ranked by number of internationally linked US trademarks.

Company Name	Total US trademarks	Internationally linked US trademarks	Design active	Contract or offshore manufacture	Owns at least one patent
ALTICOR INC.	600	304	yes	yes	yes
PRIMESOURCE BUILDING PRODUCTS INC.	352	194	not clear	yes	yes
SYSCO CORP	517	163	yes	yes	yes
SCHOLASTIC INC.	1103	135	yes	yes	yes
ALIPHCOM	270	133	yes	not clear	yes
GOOGLE INC.	388	131	yes	yes	yes
FERGUSON ENTERPRISES INC.	343	104	yes	yes	yes
POMWONDERFUL LLC	271	95	yes	yes	yes
TRUE VALUE COMPANY	781	95	yes	yes	no
TELEBRANDS CORP.	370	89	yes	yes	yes

CONCLUDING COMMENTS

Technology has ushered in new manufacturing methods that have permanently altered the distribution and organisation of global production and how the subsequent revenue streams are collected. Value is no longer created and controlled at the point of manufacture. Anecdotally, we know that revenue for select firms and industries can accrue at the upstream (R&D) or downstream (sales and service) ends of the production chain, yet we have limited systematic evidence for the bulk of firms and industries.

In this article, we advance a novel metric of the global footprint of US companies that leverages the 'paper trail' associated with businesses that are (prospectively) active in overseas markets. There is a strong commercial imperative for companies desiring to launch a new product to obtain their own trademark for that market.

We develop and describe TM-Link – a worldwide trademark database that incorporates a neural network linking algorithm to identify equivalent international trademark applications, filed in different legal jurisdictions by the same company to protect the same product or service. This exercise reveals that about 12 per cent of US trademarks filed by US domiciled firms have an equivalent in one of the other five jurisdictions.

To evaluate the relationship between international trademarks and the global footprint of US firms the data is aggregated to 4-digit NAICS industry across the 50 States from 2003 through 2018. As predicted, the share of international trademarking is positively associated with both exports and outward FDI. We show that the remaining variation in international trademarking activity (after controlling for outward FDI and exporting) conforms to known patterns of arms-length offshoring (by state and by industry).

Perhaps our most striking result is that variation in international trademarking activity, after controlling for exports and outward FDI, maps closely to product complexity measures such as those developed in Felipe et al. (2012). As complex final goods are likely to contain numerous parts and components, it is reasonable that firms within these industries have more opportunities to deal with overseas suppliers in at least some proportion of their supply chains.

To the extent this fabless mode of production is significant, relying on export and FDI measures to calculate comparative industry advantage or industrial capabilities of regions or nations will be inaccurate. Trademarks also offer an opportunity to understand innovations in service industries, a limitation that has often hampered research in patents. Finally, we demonstrate micro-geographical potential of international trademark data. Our linked database of international trademark equivalents is free to use by other scholars and we believe it will yield further insights across international business, industrial organization and economic geography.

ACKNOWLEDGEMENTS

We would like to thank the participants at the Asia Pacific Innovation Conference, Beijing, China; 10-11 October 2019, 5th Geography of Innovation Conference, Stavanger, Norway; 29-31 January 2020, seminar participants at the United States Patent and Trademark Office, Copenhagen Business School and University of Liverpool, for helpful comments. Funder the Australian Research Council and IP Australia. Grant LP160100776.

REFERENCES

Athukorala, P.C. and Yamashita, N., (2006). Production fragmentation and trade integration: East Asia in a global context. *The North American Journal of Economics and Finance*, *17*(3), pp.233-256.

Baldwin, R. and Venables, A.J., (2013). Spiders and snakes: Offshoring and agglomeration in the global economy. *Journal of International Economics*, *90*(2), pp.245-254.

Baroncelli, E., Fink, C., and Javorcik, B. (2005). The Global Distribution of Trademarks: Some Stylised Facts' *World Economy*, 28, 765-782

Bayard, K., Byrne, D. and Smith, D. (2015) 'The Scope of U.S. "Factoryless Manufacturing"', Chapter 4 (pp. 81-120) in: Measuring Globalization: Better Trade Statistics for Better Policy, Volume 2, Factoryless Manufacturing, Global Supply Chains, and Trade in Intangibles and Data Susan N. Houseman and Michael Mandel, eds. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research, 2015.

Bernard, A.B. and Fort, T.C., (2015). Factoryless goods producing firms. *American Economic Review*, *105*(5), 518-23.

Bettencourt, L.M., Lobo, J. and Strumsky, D., (2007). Invention in the city: Increasing returns to patenting as a scaling function of metropolitan size. *Research policy*, 36(1), pp.107-120.

Block, J. H., Fisch, C. O., Hahn, A. and Sandner, P. G. (2015). Why do SMEs file trademarks? Insights from firms in innovative industries. *Research Policy*, 44, 1915-1930.

Castaldi C. (2020). All the great things you can do with trademark data: Taking stock and looking ahead. *Strategic Organization*, 18, 472-484.

Caves, R. E. (1974). 'Causes of direct investment: Foreign firms' shares in Canadian and United Kingdom manufacturing industries', *The Review of Economics and Statistics* 56(3), 279–293.

De Faria, P. and Sofka, W., (2010). Knowledge protection strategies of multinational firms—A cross-country comparison. *Research Policy*, *39*(7), 956-968.

Delgado, M. (2020). The co-location of innovation and production in clusters. *Industry and Innovation*, 27(8), 842-870.

Department of Energy (2005). U.S. Light Vehicle Manufacturing Locations, 2004. http://web.archive.org/web/20120207003830/http://www1.eere.energy.gov/vehiclesandfuels/facts/2005/fcvt_fotw383.html. Last Accessed: 16 January 2021.

Dinlersoz, E., Goldschlag, N., Myers, A. and Zolas, N., (2018). An anatomy of US firms seeking trademark registration. In *Measuring and Accounting for Innovation in the 21st Century*. University of Chicago Press.

Dissart, J.C., (2003). Regional economic diversity and regional economic stability: research results and agenda. *International Regional Science Review*, *26*(4), 423-446.

Doherty, M. (2015) 'Reflecting Factoryless Goods production in the U.S. Statistical System', Chapter 2 (pp. 13-44) in: Measuring Globalization: Better Trade Statistics for Better Policy, Volume 2, Factoryless Manufacturing, Global Supply Chains, and Trade in Intangibles and Data Susan N. Houseman and Michael Mandel, eds. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research, 2015.

Dowd, T, Landefeld, P., and Moore, A. (2017). Profit shifting of U.S. multinationals, *Journal of Public Economics*, 148, 1–13.

Dunning, J. H. (1979). 'Explaining changing patterns of international production: in defence of the eclectic theory', Oxford Bulletin of Economics and Statistics 41(4), 269–295.

Economides, N. (1996). The economics of networks. *International Journal of Industrial organization*, 14, 673-699.

Etemad, H., (2004). Internationalization of small and medium-sized enterprises: a grounded theoretical framework and an overview. *Canadian Journal of Administrative Sciences*, 21(1), 1.

Felipe, J., Kumar, U., Abdon, A., and Bacate, M. (2012). Product complexity and economic development. *Structural Change and Economic Dynamics*, 23, 36-68.

Fink, C., Javorcik, B.S. and Spatareanu, M., (2005). Income-related biases in international trade: what do trademark registration data tell us?. *Review of World Economics*, 141, 79-103.

Flikkema M. J., Castaldi, C., de Man A. P. *et al.* (2019). Trademark's relatedness to product and service innovation: A branding strategy approach. *Research Policy* vol. 48, pp. 1340–1353.

Flikkema, M. J., de Man, A. P. and Castaldi, C. (2014). Are trademark counts a valid indicator of innovation? Results of an in-depth study of new Benelux trademarks filed by SMEs. *Industry and Innovation*, vol. 21, pp. 310–331.

Fort, T., Pierce, J., and Schott, P. (2018). New perspectives on the decline of US manufacturing employment. *Journal of Economic Perspectives*, 32(2), 47-72.

Fosfuri, A. and Giarratana, M.S., (2009). Masters of war: Rivals' product innovation and new advertising in mature product markets. *Management Science*, *55*(2), 181-191.

Fosfuri, A., Giarratana, M., and Luzzi, A. (2008). The Penguin Has Entered the Building: The Commercialization of Open Source Software Products. *Organizational Science, 19,* 292-305

Gereffi, G., Humphrey, J. and Sturgeon, T., (2005). The governance of global value chains. *Review of International Political Economy*, 12(1), 78-104.

Giarratana, M. S. and Torrisi, S. (2010). Foreign entry and survival in a knowledge-intensive market: Emerging economy countries' international linkages, technology competences, and firm experience. *Strategic Entrepreneurship Journal*, 4, 85–104.

Gotsch, M. and Hipp, C. (2014). Using trademarks to measure innovation in knowledge-intensive business services. *Technology Innovation Management Review*, 4, 18.

Graham, S.J., Hancock, G., Marco, A.C. and Myers, A.F., (2013). The USPTO trademark case files dataset: Descriptions, lessons, and insights. Journal of Economics & Management Strategy, 22(4), 669-705.

Greenhalgh, C., Longland, M. and Bosworth, D., (2003). Trends and distribution of intellectual property: UK and European patents and UK trade and service marks 1986-2000'. *Report to the UK Patent Office on a Research Project into The Extent and Value of Intellectual Property in United Kingdom Firms*.

Grubert, H. (2012). Foreign Taxes and The Growing Share of U.S. Multinational Company Income Abroad: Profits, Not Sales, Are Being Globalized, *National Tax Journal*, 65, 247–282.

Herz, B., and Thomson, R., (2021) Retail Foreign Direct Investment by Brand Manufacturers, CTI Working Paper

Hymer, S.H., (1976). International operations of national firms. MIT press.

Jensen, P. and Webster, E. (2004). 'Recent patterns of trade marking activity in Australia,' *Australian Intellectual Property Journal*, 15(2), 112-126.

Johnson, R.C., (2018). Measuring global value chains. *Annual Review of Economics*, *10*, pp.207-236.

Jones, R.W. and Kierzkowski, H., (2000). A framework for fragmentation (No. 00-056/2). Tinbergen Institute Discussion Paper.

Kim K., Khabsa, M. and Giles, C. L. (2016). Random Forest DBSCAN for USPTO Inventor Name Disambiguation. arXiv:1602.01792, [cs.IR]

Klier, T. and Rubenstein, J. (2008). Who Really Made Your Car? Restructuring and Geographic Change in the Auto Industry. *W.E. Upjohn Institute for Employment Research*, Kalamazoo, MI.

Knight, G, and Liesch, P. (2016). Internationalization: From incremental to born global, *Journal of World Business*, 51, 93–102.

Krasnikov, A., Mishra, S. and Orozco, D., (2009). Evaluating the financial impact of branding using trademarks: A framework and empirical evidence. *Journal of Marketing*, 73(6), 154-166.

Krizhevsky, A., Sutskever, I. and Hinton, G. E., (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, 25,1097–1105.

Lancaster, K., (1990). The economics of product variety: A survey. *Marketing science*, 9(3), pp.189-206.

Li, G. C., Lai, R., D'Amour, A., Doolin, D. M., Sun, Y., Torvik, V.I., Yu, A. Z., and Lee, F. (2014). Disambiguation and co-authorship networks of the U.S. patent inventor database (1975-2010). *Research Policy*, 43, 941–955.

Malmberg, C., (2005). 'Trademark Statistics as Innovation Indicators? - A Micro Study', CIRCLE Electronic Working Paper Series, 2005-17, CIRCLE (Centre for Innovation, Research and Competence in the Learning Economy), Lund University

Mendonça, S., Pereira, T. S. and Godinho, M. M. (2004). 'Trademarks as an Indicator of Innovation and Industrial Change', *Research Policy*, 33, 1385-1404.

Morikawa, M., (2016). Factoryless goods producers in Japan. *Japan and the World Economy*, *40*, 9-15.

Morrison, G., Riccaboni, M. and Pammolli, F. (2017). Disambiguation of patent inventors and assignees using high-resolution geolocation data. *Scientific Data*, 4, 1–21.

Nasirov, S. (2018). The Use of Trademarks in Empirical Research: Towards an Integrated Framework. *Available at SSRN 3296064*.

Neubig, T., and Wunsch-Vincent, S. (2017). A missing link in the analysis of global value chains: cross-border flows of intangible assets, taxation and related measurement implications, WIPO Economic Research Working Paper No. 37

Nicol, D., Liddicoat, J. and Critchley, C., (2013). A role for virtual biotechnology companies in drug discovery and development*? Journal of Commercial Biotechnology*. 19(3), 10-19.

Petrie, S. and Julius, T. (2019). Representing text as abstract images enables image classifiers to also simultaneously classify text, arXiv:1908.07846, [cs.CL].

Petrie, S., Adams, M., Mitra-Kahn, B., Johnson, M., Thomson, R., Jensen, P., Palangkaraya, A. and Webster, E. (2020). TM-Link: An Internationally Linked Trademark Database, *The Australian Economic Review*, vol. 53, no. 2, pp. 254–269.

Porter, M.E. (1998). Clusters and competition: new agendas for companies, governments, and institutions. Published in *On Competition*, Porter, M.E., 197-287. Harvard Business School Press, Boston, MA.

Ramello, G.B. (2006). What's in a Sign? Trademark Law and Economic Theory. *Journal of Economic Surveys*, 20, 547-565.

Ramello, G.B. and Silva, F., (2006). Appropriating signs and meaning: the elusive economics of trademark. *Industrial and Corporate Change*, 15, 937-963.

Rialp, A., Rialp, J. and Knight, G.A., (2005). The phenomenon of early internationalizing firms: what do we know after a decade (1993–2003) of scientific inquiry? *International Business Review*, 14(2), 147-166.

Sandner, P. and Block, J. (2011). The market value of R&D, patents, and trademarks, *Research Policy*, 40, 969-983.

Schautschick, P. and Greenhalgh, C., (2016). Empirical studies of trade marks–the existing economic literature. *Economics of innovation and new technology*, *25*(4), 358-390.

Schmoch, U. (2003). 'Service Marks as Novel Innovation Indicator', *Research Evaluation*, vol. 12.

Schmoch, U. and Gauch, S. (2009). Service marks as indicators for innovation in knowledge-based services. *Research Evaluation*, 18, 323–335

Schmoch, U., (2014). Knowledge transfer from German universities into the service sector as reflected by service marks. *Research Evaluation*, *23*(4), 341-351.

Stoneman, P. (2010). Soft Innovation: Economics, Product Aesthetics, and the Creative Industries. Oxford: Oxford University Press.

Styles, C., Loane, S. and Bell, J., (2006). Rapid internationalisation among entrepreneurial firms in Australia, Canada, Ireland and New Zealand. *International Marketing Review*. 23, 467-485

Tomiura, E., (2018). *Cross-Border Outsourcing and Boundaries of Japanese Firms. Advances in Japanese Business and Economics*. Volume 18, Springer, Springer Nature Singapore Pte Ltd

Ventura, S. L. R. Nugent, and E. R. H. Fuchs. (2015). Seeing the non-stars: (Some) sources of bias in past disambiguation approaches and a new public tool leveraging labeled records. *Research Policy*, 44,1672–1701.

WIPO (2004). WIPO Intellectual Property Handbook, World Intellectual Property Office, Geneva.

WIPO (2017). World Intellectual Property Report 2017 Intangible Capital in Global Value Chains, World Intellectual Property Office, Geneva.

Xing, Y. (2017). Global Value Chains and the Missing Exports of the United States, GRIPS Discussion Paper 17-06

Xing, Y. and N. Detert. (2010). How the iPhone Widens the United States Trade Deficit with the People's Republic of China. ADBI Working Paper 257. Tokyo: Asian Development Bank Institute

Xing, Y., (2020). Global value chains and the “missing exports” of the United States. *China Economic Review*, 101429.

Yang G.-C., Liang, C., Jing, Z., Wang, D.-R., and Zhang, H.-C. (2017). A Mixture Record Linkage Approach for US Patent Inventor Disambiguation. Advanced Multimedia and Ubiquitous Engineering. FutureTech 2017, MUE 2017. Lecture Notes in Electrical Engineering, 448, 331–338

APPENDIX A – LINKING ALGORITHM

Our linking algorithm involves two main facets: a binning algorithm and neural network classification algorithm. The binning algorithm groups similar trademark applications with shared (exact matched) strings derived from trademark text. The process involves systematically simplifying trademark text (e.g., removing punctuation).[19] This avoids the issue that exact matching the original trademark text would miss many equivalent trademarks because of minor text differences. Applications which do not share any Nice classification numbers with other applications in the same bin are removed. The procedure aims to maximize the collection of similar trademarks into the same bin by increasing their consistency, while avoiding the generation of large, non-specific bins which would contain many false positive associations.

To further reduce the proportion of false positives, we trained a machine learning algorithm to identify and remove as many false positive links as possible. The neural network machine learning algorithm is based on an adaptation of Petrie & Julius (2019).[20] The algorithm processes a single within-bin pair of TM applications at a time, classifying the TM-TM pair as either matched (equivalent TMs) or non-matched (not equivalent TMs). This is done by converting text-based trademark application data into a stacked 2D color image (or, equivalently, a 3D tensor) representation of that text.[21] We then train a modified version of the "AlexNet" neural network (Krizhevsky et al. 2012) to perform the pairwise match/non-match classification, as described in Petrie & Julius (2018).

Figure A1 illustrates how we convert text to abstract image for the word "JEN" (rightmost image) (Petrie & Julius 2019). To generate this image, we firstly define a specific grid layout of alphabet characters (this grid is present in each of the images in Figure A1). We then add a particular color (e.g. red) to pixels corresponding

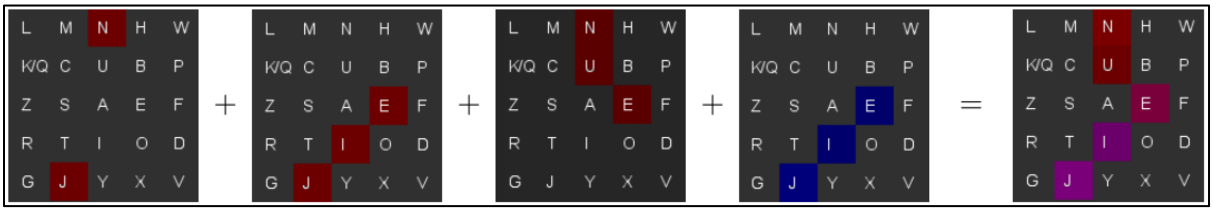
[19] Punctuation characters were removed if word mark text length is greater than five characters (the five-character limit avoids removal of useful punctuation from short text strings). All instances of "AND" or "+" are replaced with "&". Since it is common for word marks to also contain acronyms (e.g. "GENERAL ELECTRIC GE") in some offices but not others, we attempt to remove such acronyms using rule-based text replacement. All characters are converted to upper case, and all single-character punctuation which is surrounded by whitespace is removed. All remaining whitespace is then removed, along with any remaining "+" characters. Blocks containing single-character word mark text (e.g. "S") are disbanded (i.e. separated into different blocks), since those blocks are highly non-specific (such blocks would lead to a large increase in the number of false positive TM-TM links without substantially increasing the number of true positive links).

[20] Petrie and Julius (2019) is focused on inventor name disambiguation and achieves F1 Score 99.09%, precision: 99.41%, recall: 98.76%, outperforming world standard (Li *et al.* 2014; Ventura *et al.* 2015; Kim *et al.* 2016; Morrison *et al.* 2017; Yang *et al.* 2017).

[21] The process for converting text to image is described in Appendix A.

to letters in the word JEN and also to pixels that connect those letters – i.e. we add red to the pixels corresponding to the first letter "J" and last letter "N" (first image), to the pixels connecting the bigram "JE" (second image), and to the pixels connecting the bigram "EN" (third image). We also add blue to the first bigram in the word (fourth image) in order to highlight the beginning of the word. The final abstract image representation of the word "JEN" is shown in the right-most image. For each within-block TM-TM pair, we use this process to convert trademark text, applicant name, and Nice classification numbers into an abstract image of each record, one in red and the other in green (with blue for the lead bigram in each case). The two images are then overlaid to represent the comparison of the TM-TM pair. Similar records have overlapping red and green pixels, which produce yellow in the RGB color scheme.

Figure A1: Illustration of process for abstract representation of text in trademark filings



Note: Approach is illustrated for the word "JEN". To convert text to an abstract image we begin defining a grid layout of alphabet characters. First and last letters are marked in a nominated color (red in this case). The color (red in this case) is then added to pixels corresponding to letters in the word JEN and also to pixels that connect those letters – i.e. we add red to the pixels corresponding to the first letter "J" and last letter "N" (first image), to the pixels connecting the bigram "JE" (second image), and to the pixels connecting the bigram "EN" (third image). We also add blue to the first bigram in the word (fourth image) in order to highlight the beginning of the word. The final abstract image representation of the word "JEN" is shown in the right-most image.

Our approach of converting text into a stacked 2D RGB bitmap for neural network-based image classification has several advantages. Most importantly, powerful classification capabilities of existing image classification networks can be utilized for text record matching. The neural network also learns its own features from the data, rather than learning from a feature vector of pre-defined string similarity measures chosen by the researcher. Since minor spelling variations and errors do not alter the resulting image very much, the neural network is capable of learning that such minor features are unimportant for discriminating between matches and non-matches. Matched records which have different word ordering are likely to be matched due to overlapping pixels in the associated abstract image representation. The neural network can also potentially learn to ignore certain shapes of common words (e.g. “Ltd”, “LLC”, “Incorporated”, etc.) which are not useful for discrimination decisions.

Our neural network linking algorithm is trained on a sub-sample of labelled TM-TM pairs, including a set of true positive links from the Madrid Protocol[22], and true negative examples from a random sample of any within-bin TM-TM pairs with very different filing dates (over 20 years), different filing jurisdictions and non-identical applicants. This labelled sub-sample is used to train the neural network to identify whether a given pairwise image comparison represents an equivalent pair of TM applications. The trained neural network can then be used to classify each within-bin TM-TM pair as either matched or non-matched.

To separate cases of multiple firms applying for the same trademark, we require that all linked trademark applications share an applicant with one or more equivalent trademarks in the same family (Petrie et al. 2020). We do this by requiring that each application's applicant name has a high degree of string similarity to the applicant name of one or more trademarks in the same family group. We remove any applications that do not meet this requirement, and assign them to a different family ID.

Our algorithms used data from the USPTO Trademark Case File Dataset (TCFD)[23], the Australian Intellectual Property Open Government Data (IPGOD), the New Zealand Trademark Information API, the Canadian Trademarks Database, the UKIPO Trade Mark Data Release and the EUIPO Community Trademark Database.[24] All of these trademark data sources capture information from applicants that are seeking trademark registration. The data sources do not include common law trademarks, which firms may rely upon in the course of trade.

[22] The Madrid Agreement Concerning the International Registration of Marks (Madrid Agreement 1891) and the Protection Relating to the Madrid Agreement Concerning the International Registration of Marks (Madrid Protocol 1989), collectively known as the Madrid System (or the International Trademark System), is an international mechanism that allows applicants to submit a single trademark application with a set of fees can designate the countries in which the application seeks registration. The system administered by WIPO allows applicants to seek protection in up to 117 countries. The intention of the Madrid System is to reduce the administrative burden on applicants filing in multiple jurisdictions and facilitates the maintenance of trademark rights abroad.

[23] We did not update the records due to any changes in ownership after application.

[24] The European Union trademark applications are only for those applications that target the entire European Union bloc, not those of individual countries.

APPENDIX B – Building NAICS-Niece Concordance

To develop a concordance table between trademark data and export data (reported in NAICS) we matched the owners of 4,114,075 trademark applications in TM-Link (representing 1,619,606 unique owners) between 1996 and 2013 to the corresponding record among the 25.6 million U.S firms listed in Orbis with at least one NAICS code. Our general approach was to preference accuracy over recall and to minimize false positive contamination. First, we undertook basic string cleaning of company names in Orbis and trademark owner name. Strings are first lowercased, common abbreviations are expanded and standard identifiers are removed. For each unique trademark applicant, we identified candidate matches based on optimizing string similarity metrics under various geographic similarity requirements. Specifically, we identified the most similar company name in Orbis (Levenshtein distance) (1) globally, (2) within the subset in Orbis in the same geography, and (3) for the subset in Orbis with shared words that are not in the standard English dictionary. Geography included: match of city name, state, postcode, contiguous postcode. We then use a decision tree approach, grouping ties according to combinations of similarity criteria and selected the best candidates for each unique USPTO_owner entry. The computational work was undertaken on Swinburne University of Technology's supercomputer (OzSTAR) using a custom Python3 pipeline.

Table B1: US domicile trademark applications matched to Orbis, 1996 to 2013

	Total		With international equivalents	
Trademark applications	4,114,075		362,041	
Matched to Orbis (percent)	2,023,029	(49.2)	225,888	(62.4)

Source. Authors' calculations

APPENDIX C – DEFINITIONS

Table C1: Variable Definitions

Variable	Definition	Source
International Trademark Applications	Stock of trademark applications originating in a state-industry group-year identified as having an equivalent trademark in an international jurisdiction.	TM-Link; Orbis
Trademark Applications	Stock of all trademark applications originating in a state-industry group-year.	TM-Link; Orbis
Share Trademark Applications with an International Link	The fraction of trademark applications in a state-industry group-year which have been linked to a trademark in the following country/regions: Australia, Canada, European Union, or New Zealand.	TM-Link; Orbis
Exports ($2012 millions)	Real value of gross exports in a state-industry group-year.	USA Trade Online Origin of Movement Series (US Census Bureau)
FDI ($2012 millions)	Real value of stock of outgoing foreign direct investment within the state-industry group-year	fDi Markets
Wholesale Establishments	Number of wholesale establishments within the state-year. This does not vary across industry groups.	Quarterly Census of Employment and Wages
Establishments	Number of establishments within the state-industry group-year.	County Business Patterns Survey

Note. Unit of analysis is State-Industry Group-Year where Industry is the 4-digit NAICS. Exports and FDI are adjusted for inflation using the Consumer Price Index for all urban consumers (CPI-U) with 2012 set as the base year.

APPENDIX D – FULL SET OF RESIDUALS BY NAICS

Table D1: Average Residual Value from Column (1) in Table 3 by Industry Group (4-Digit NAICS)

Rank	No	Industry Group (NAICS)	Residual
1	3369	Other Transportation Equipment Manufacturing	4.4
2	3362	Motor Vehicle Body and Trailer Manufacturing	4.2
3	3351	Electric Lighting Equipment Manufacturing	3.9
4	3334	Ventilation, Heating, Air-Conditioning, and Commercial Refrigeration Equipment Manufacturing	3.4
5	4231	Motor Vehicle and Motor Vehicle Parts and Supplies Merchant Wholesalers	2.9
6	3365	Railroad Rolling Stock Manufacturing	2.9
7	3366	Ship and Boat Building	2.8
8	3352	Household Appliance Manufacturing	2.7
9	3335	Metalworking Machinery Manufacturing	2.7
10	3363	Motor Vehicle Parts Manufacturing	2.5
11	3321	Forging and Stamping	2.3
12	4246	Chemical and Allied Products Merchant Wholesalers	2.2
13	4242	Drugs and Druggists' Sundries Merchant Wholesalers	2.2
14	3326	Spring and Wire Product Manufacturing	2.1
15	3315	Foundries	2.1
16	3262	Rubber Product Manufacturing	2.1
17	4237	Hardware, and Plumbing and Heating Equipment and Supplies Merchant Wholesalers	2.0
18	3336	Engine, Turbine, and Power Transmission Equipment Manufacturing	2.0
19	4238	Machinery, Equipment, and Supplies Merchant Wholesalers	2.0
20	3331	Agriculture, Construction, and Mining Machinery Manufacturing	1.9
21	3339	Other General Purpose Machinery Manufacturing	1.8
22	3325	Hardware Manufacturing	1.8
23	3353	Electrical Equipment Manufacturing	1.8
24	4236	Electrical and Electronic Goods Merchant Wholesalers	1.7
25	3279	Other Nonmetallic Mineral Product Manufacturing	1.7
26	3322	Cutlery and Handtool Manufacturing	1.5
27	3312	Steel Product Manufacturing from Purchased Steel	1.4
28	3253	Pesticide, Fertilizer, and Other Agricultural Chemical Manufacturing	1.4
29	4234	Professional and Commercial Equipment and Supplies Merchant Wholesalers	1.4
30	3343	Audio and Video Equipment Manufacturing	1.2
31	3332	Industrial Machinery Manufacturing	1.1
32	3361	Motor Vehicle Manufacturing	1.1
33	3327	Machine Shops; Turned Product; and Screw, Nut, and Bolt Manufacturing	1.0
34	3333	Commercial and Service Industry Machinery Manufacturing	1.0
35	3273	Cement and Concrete Product Manufacturing	1.0
36	3271	Clay Product and Refractory Manufacturing	1.0
37	3391	Medical Equipment and Supplies Manufacturing	0.8
38	3323	Architectural and Structural Metals Manufacturing	0.7
39	3324	Boiler, Tank, and Shipping Container Manufacturing	0.6
40	4233	Lumber and Other Construction Materials Merchant Wholesalers	0.6
41	3256	Soap, Cleaning Compound, and Toilet Preparation Manufacturing	0.6
42	3255	Paint, Coating, and Adhesive Manufacturing	0.5

43	3212	Veneer, Plywood, and Engineered Wood Product Manufacturing	0.5
44	4239	Miscellaneous Durable Goods Merchant Wholesalers	0.3
45	3122	Tobacco Manufacturing	0.3
46	3313	Alumina and Aluminum Production and Processing	0.2
47	4249	Miscellaneous Nondurable Goods Merchant Wholesalers	0.1
48	3379	Other Furniture Related Product Manufacturing	0.0
49	3259	Other Chemical Product and Preparation Manufacturing	0.0
50	3252	Resin, Synthetic Rubber, and Artificial Synthetic Fibers and Filaments Manufacturing	0.0
51	3314	Nonferrous Metal (except Aluminum) Production and Processing	-0.1
52	3359	Other Electrical Equipment and Component Manufacturing	-0.1
53	4235	Metal and Mineral (except Petroleum) Merchant Wholesalers	-0.2
54	3329	Other Fabricated Metal Product Manufacturing	-0.2
55	3272	Glass and Glass Product Manufacturing	-0.3
56	3219	Other Wood Product Manufacturing	-0.3
57	3311	Iron and Steel Mills and Ferroalloy Manufacturing	-0.3
58	4232	Furniture and Home Furnishing Merchant Wholesalers	-0.3
59	3372	Office Furniture (including Fixtures) Manufacturing	-0.4
60	3149	Other Textile Product Mills	-0.4
61	3345	Navigational, Measuring, Electromedical, and Control Instruments Manufacturing	-0.4
62	3211	Sawmills and Wood Preservation	-0.5
63	3111	Animal Food Manufacturing	-0.5
64	4248	Beer, Wine, and Distilled Alcoholic Beverage Merchant Wholesalers	-0.8
65	3344	Semiconductor and Other Electronic Component Manufacturing	-0.8
66	3371	Household and Institutional Furniture and Kitchen Cabinet Manufacturing	-0.8
67	3342	Communications Equipment Manufacturing	-0.9
68	3131	Fiber, Yarn, and Thread Mills	-0.9
69	3169	Other Leather and Allied Product Manufacturing	-0.9
70	3133	Textile and Fabric Finishing and Fabric Coating Mills	-1.0
71	3346	Manufacturing and Reproducing Magnetic and Optical Media	-1.1
72	3141	Textile Furnishings Mills	-1.1
73	4243	Apparel, Piece Goods, and Notions Merchant Wholesalers	-1.2
74	4241	Paper and Paper Product Merchant Wholesalers	-1.3
75	4245	Farm Product Raw Material Merchant Wholesalers	-1.4
76	3251	Basic Chemical Manufacturing	-1.4
77	3161	Leather and Hide Tanning and Finishing	-1.4
78	4244	Grocery and Related Product Merchant Wholesalers	-1.4
79	3254	Pharmaceutical and Medicine Manufacturing	-1.4
80	3341	Computer and Peripheral Equipment Manufacturing	-1.5
81	3261	Plastics Product Manufacturing	-1.5
82	3159	Apparel Accessories and Other Apparel Manufacturing	-1.6
83	3364	Aerospace Product and Parts Manufacturing	-1.7
84	4247	Petroleum and Petroleum Products Merchant Wholesalers	-1.8
85	3132	Fabric Mills	-1.9
86	3117	Seafood Product Preparation and Packaging	-1.9
87	3162	Footwear Manufacturing	-2.0
88	3399	Other Miscellaneous Manufacturing	-2.0

89	3241	Petroleum and Coal Products Manufacturing	-2.1
90	3152	Cut and Sew Apparel Manufacturing	-2.3
91	3222	Converted Paper Product Manufacturing	-2.3
92	3121	Beverage Manufacturing	-2.4
93	3113	Sugar and Confectionery Product Manufacturing	-2.6
94	3115	Dairy Product Manufacturing	-2.6
95	4251	Wholesale Electronic Markets and Agents and Brokers	-2.8
96	3114	Fruit and Vegetable Preserving and Specialty Food Manufacturing	-3.2
97	3112	Grain and Oilseed Milling	-3.2
98	3119	Other Food Manufacturing	-3.4
99	3221	Pulp, Paper, and Paperboard Mills	-3.4
100	3118	Bakeries and Tortilla Manufacturing	-3.4
101	3231	Printing and Related Support Activities	-4.0
102	3116	Animal Slaughtering and Processing	-4.4

Note. The Residual column indicates the average value of the residuals within a 4-digit NAICS industry derived from Model (3) in the primary regression results in Table 4. See Appendix C for variable definitions. Rankings are robust to the specification.