# An Anatomy of Trademarking by Firms in the United States[*]

Emin M. Dinlersoz[†]  Nathan Goldschlag[‡]  Amanda Myers[§]  Nikolas Zolas[¶]

March 2017

**\* Preliminary and incomplete \***

## Abstract

This paper reports on the construction of a new dataset that combines data on trademarks from the U.S. Patent and Trademark Office with the microdata on firms from the U.S. Census Bureau. The methodology for merging the data and identifying matches between trademarks and firms is described. The resulting dataset allows tracking of various activity related to trademarks over the life-cycle of firms, such as the first application for a trademark, the first use of a trademark, and the renewal, assignment, and abandonment of trademarks. Some facts about firm-level trademarking activity are documented, including the incidence and timing of trademarking activity over the firm life-cycle, the connection between firm characteristics and trademarking, and the relation of trademarks to firm growth. The dataset offers new possibilities for research on how trademarks are related to firm dynamics and performance, firm-level innovative activity and product introductions, and firm strategies aimed at acquiring customers, generating loyalty, building brands and reputation, and signalling quality.

[†]Center for Economic Studies, U.S. Census Bureau. E-mail: `emin.m.dinlersoz@census.gov`.

[‡]Center for Economic Studies, U.S. Census Bureau. E-mail: `nathan.goldschlag@census.gov`

[§]U.S. Patent and Trademark Office. E-mail: `amanda.meyers@uspto.gov`

[¶]Corresponding Author. Center for Economic Studies, U.S. Census Bureau, 4600 Silver Hill Road, Suitland, MD 20746. E-mail: `nikolas.j.zolas@census.gov`

# 1  Introduction

Attracting consumers and generating loyalty is an important objective for any firm that seeks to survive and grow. Among many ways to achieve this objective, trademarks are a peculiar one. By protecting a firm's intangible assets, trademarks make it possible for firms to reduce consumer search and switching costs, lower the cost of introducing and marketing new products, and generate brand awareness and loyalty. While trademarks are in general aimed at protecting goodwill for a firm's products and services, they may not always be directly linked to a particular attribute of the firm that generates this goodwill. Goodwill can be generated through consistent quality, a distinctive set of products, or a unique service that make a firm stand out. The fact that relatively few firms utilize trademarks suggests that certain firms stand to gain more from trademarking than others.

Empirical analysis of trademarking by U.S. firms has been limited mainly because of a lack of comprehensive data on firm-level trademark activity in the United States. Prior empirical research on trademarks has relied mostly on data from Europe, particularly the U.K. and France, and from Australia. Recently, substantial progress has been made on this front in the United States. The United States Patent and Trademark Office (USPTO) has made available trademark data that covers 6.7 million trademark applications for the period 1870-2015. The USPTO Trademark Case Files Dataset is a remarkable synthesis of various activities conducted by firms regarding trademarking.[1] It contains detailed information on applications for, as well as registrations, commercial use, renewals, assignments, cancellations, and abandonments of trademarks. It can be used to identify the timing of filings and registrations of trademarks, how many trademarks a firm holds, and in what categories or classes of goods and services the trademarks fall. It also has data on oppositions filed against a trademark and resulting prosecution, which may contain useful information on the value of a trademark. Despite the wealth of information contained in this dataset, it says little about the firms that take actions related to trademarks, and when in their life-cycle they take these actions. As such, it is not possible to directly use the USPTO Case Files Dataset to uncover how trademarking activity is related to firm characteristics and performance. It is therefore important to combine the trademark data with firm characteristics to understand which types of firms trademark and when, and the causes and consequences of trademarking.

This paper reports on the construction of a new database that combines USPTO's trademark data with firm-level micro data at U.S. Census Bureau. This new dataset fills the void in connecting trademark activity with firm characteristics, performance, and dynamics in the United

---

[1]See Graham, Hancock, Marco and Myers (2016) for details on the construction and features of the USPTO Trademark Case Files Dataset.

States. The dataset provides information on the timing and incidence of trademarks over the life-cycle of a firm by using the longitudinal nature of the firm-level data from the U.S. Census Bureau. As a result, the dataset can be used to explore not only what kind of firms trademark, but also when in their life-cycle they trademark, and how trademarking is related to key firm outcomes, such as entry, survival, and employment or revenue growth. Some exercises are carried out with the new dataset to demonstrate the relationship between trademarks and firm characteristics.

The dataset opens several research possibilities regarding key events in the life-cycle of a firm that triggers the emergence of highly successful firms and leads to skewness in firm outcomes. There is a growing interest in identifying early precursors of successful businesses. For instance, recent research finds that having a patent or a trademark application is highly correlated with the ultimate success of early entrepreneurial activity, where success is measured by rare events such as an IPO or a high-value acquisition.[2] Similarly, patenting and R&D activity are found to be positively correlated with measures of firm growth.[3] The linked firm-trademark dataset allows for exploring the connection between various trademark activity and continuous measures of firm performance, such as employment and revenue growth, beyond what is captured in discrete firm outcomes, such as IPO or entry. The longitudinal nature of the data can also help with identifying whether trademarks are lagging or leading indicators of firm performance.

Another avenue of research that can benefit from the new dataset is the research in the valuation of trademarks. The dataset allows for the observation of first trademarking activity, as well as subsequent trademarks, and trademark assignments. Reputation is a valuable asset that needs to be protected and that can be traded. The information contained in the trademark filings and assignments can be used, in conjunction with measures of firm value, to attach valuations to trademarks. In addition, the oppositions placed against a trademark filing and the resulting prosecutions observed in the linked dataset can also be used in an effort to identify the inherent value of a trademark, as such oppositions would be unlikely if the owner of a firm placed little or no value on a trademark.

There is also a large body of work in the marketing literature that could benefit from the data constructed here. Some theories of trademarks emphasize a connection between trademarks and customer acquisition and loyalty-building by firms. The theoretical models emphasize the role of trademarks in reducing consumer search and switching costs, establishing brand loyalty and goodwill, and signalling quality.[4] Furthermore, by protecting a firm's investments in marketing, trademarks can induce advertising and marketing expenditures, leading to changes in a firm's competitive position in a market and a change in the industry concentration. The new dataset

---

[2] See, for instance, Fazio, Guzman, Murray and Stern (2016).
[3] See, e.g., Foster, Grim and Zolas (2016).
[4] See, e.g., Landes and Posner (1987,1988) and Economides (1988).

can help in testing some of these models in conjunction with data on firms' marketing activity.

Finally, the dataset can provide valuable information on how trademarks capture innovative activity at the firm level. Other data, such as firm-level data on patents and R&D activity, can be used to correlate innovative activity with trademarks. An open question is whether trademarks are a lagging or leading indicators of innovation. If trademarks tend to follow escalated R&D and patenting activity or product introductions at the firm level, one may use trademarks as lagging indicators of innovation. If, on the other hand, patents or R&D typically follow a trademark application or registration, trademarks can be used as leading indicators of innovative activity. In addition, trademarking may signal innovations of the type not captured by patents and R&D measures, such as innovations in services. Future research plans to establish the connections between trademarks on the one hand, and patenting, R&D, and product introductions, on the other.

The rest of the paper is organized as follows. The next section gives a brief overview of the prior literature on the use of trademark data in conjunction with firm level data in other countries. Section 3 discusses the data inputs and the algorithm used to match the trademark data with firm-level data. The analysis in Section 4 documents some characteristics of firms that trademark, and provides a first look at how trademark activity is correlated with firm growth and innovative activity. Section 5 concludes.

## 2 Prior literature

There is a small but growing literature on empirical work concerning trademarks. Much of the recent work that connects firm-level data with trademark data has been limited to the context of Europe and Australia. For the United States, some research has been done using datasets such as Compustat, which covers only publicly traded firms. The lack of a comprehensive dataset that relates trademarks to the characteristics of firms that own them has been a major impediment to progress on empirical research on trademarks.

A survey by Schautshick and Greenhalgh (2013) provides an excellent summary of the current state of the empirical research on trademarks. Several conclusions emerge from their observations across the literature. First, there has been an upsurge in trademark activity since 1975 in Australia, the United Kingdom and the United States, with similar trends in time series for these countries over the period 1975-2002.[5] The services sector, as well as deregulated and restructured industries, exhibit the highest rates of growth in trademarks during this period. Interestingly, there is a similar upsurge in patents in developed counties, and the upsurge in trademarks appear to *lead* the upsurge in patents by about ten years.

---

[5]See Figure 2 in Schautschick and Greenhalgh (2013).

Second, at the country level, an increase in demand for product variety and quality appears to better explain the growth in trademarks compared to the mere expansion of output. In general, the studies surveyed indicate that trademark use is correlated with product differentiation, marketing, and innovation. However, the degree of this correlation depends on the industries investigated.

Third, studies at the firm level tend to indicate that trademarks are used by firms to protect identity and reputation, and the private value of trademarks to firms is generally positive. There is no clear conclusion yet on the social value of trademarking, which depends on how potentially wasteful expenditure by firms on building and protecting reputation sizes up with efficiency-improving and pro-competitive effects of trademarks. Further work is needed to sort out the magnitudes of these different effects.

At a broad level, Baroncelli, Fink and Javorcik (2005) document a number of regularities in the distribution of trademarks across countries. Using a panel of countries over the period 1994-1998 based on the data published by World Intellectual Property Organization, they find evidence that higher development is correlated with a dominance of domestic brands at home and a stronger presence of these brands in foreign markets, as indicated by the foreign residents' share of trademark registrations. However, they also note an increase in foreign trademark filings by developing countries, potentially reflecting the increasing exports from such countries to more developed ones, and the resulting need to protect growing brands. At the sector and industry level, they note that most trademarking occurs in industries characterized as R&D intensive, such as scientific equipment and pharmaceutical sectors. In addition, advertising-intensive industries, such as clothing, footwear, and food products, also exhibit higher intensity of trademarking. Furthermore, business services appears to be one category of industries that rank high in the intensity of trademarking – a finding also echoed by Millot (2011) in the case of Germany and France. Service marks have also shown high growth between 1980s and 2000s in developed countries, as documented by Greenhalgh, Longland, and Bosworth (2003) for United Kingdom, by Jensen and Webster (2004) for Australia, and by Graham, Hancock, Marco, and Myers (2016) for the United States. In particular, the latter study shows that service marks filings grew by nearly 50% between 1998-2000 in the United States, potentially as a result of the dot.com boom of the late 1990s.[6] These patterns clearly reflect the growing importance of the service sector in developed countries.

A handful of studies with firm-trademark matched data lead to some conclusions regarding the connection between firm characteristics and trademarking activity. For the United Kingdom, a series of papers examine trademarking activity at the firm level. Greenhalgh, Longland, and Bosworth (2003) use panel data on medium and large manufacturing firms, and find that firm size,

---

[6]See Figure 16 in Graham, Hancock, Marco and Myers (2016).

measured either by employment or sales, had a negative relation with intensity of trademarking. Smaller firms in the panel exhibited higher trademark intensity. However, the panel used in this study consists mainly of publicly listed firms with many subsidiaries, and the applicability of the results to the broader population of firms remains an open question.

Greenhalgh and Rogers (2008) use both manufacturing and service firms to investigate the role of firm characteristics on activities related to intellectual property, including trademarks. They find that, while intellectual property assets are not always monotonically related to firm size, size is nevertheless a strong predictor of whether a firm engages in activity related to trademarking and patenting. Their study also indicates diminishing returns to firm size in terms of such activity. For both services and manufacturing, the intensity of intellectual property activity related to trademarks and patents declines as firms get larger.

Sandner (2009) provides a detailed investigation of companies' trademark portfolios using the world's largest publicly traded companies. These are the companies included in the financial databases Reuters and Compustat. Sandner identifies 4,085 companies that satisfy his selection criterion of 400 million Euros or more in revenues (as of the time of his analysis) in their latest income statement. He finds that companies indeed build trademark portfolios that are aimed to deliberately protect a company's brand. He infers brand management decisions using trademarks, and finds that new product introductions call for brand management decisions regarding whether to use existing trademarks to cover the new products or apply for new trademarks. Therefore, trademark filings can reflect both the creation of new brands and the expansion of existing brands to include new products.

Sandner and Block (2011) use data from a variety of countries, including the United States, to assess the market value of trademarks, along with R&D and patents. They find a positive effect of trademarks on firm value, controlling for patenting and R&D activity by firms. The trademark data in their study comes from the Community Trademarks register provided by the Office for Harmonization in the Internal Market. Their data for the firms is mainly from the Compustat database, which contains only the publicly traded, relatively large firms. Their final dataset for the analysis consists of a relatively small set of 1,216 publicly traded firms.

One drawback of the studies at the firm-level is that they focus mainly on large publicly-traded firms, with little information on trademarking activity in small or medium firms, or in privately-held ones. On this front, Rogers et al. (2007) build a database that spans the entire set of firms in the United Kingdom for the period 2001-2005. This database contains millions of small to medium sized firms matched with trademarks, and substantially expands the scope of the studies discussed earlier, even though the time period covered is short. One main conclusion that emerges from this study is that small firms tend to have, in proportion to their assets, higher intensity of intellectual property compared to larger firms. This result encourages further

scrutiny into relatively small and medium firms' trademarking activity.

The existing literature summarized so far points to a lack of a comprehensive longitudinally linked firm-level trademark data in the United States, especially for privately-held firms. Most research, with a few exceptions, has focused on publicly-traded, large firms. The rest of the paper describes how a more comprehensive dataset is constructed for the United States by bringing together the set of all trademark related activity available in the USPTO Case Files Dataset with the universe of all firms, public and private, available in the micro data at the U.S. Census Bureau. It then carries out a few exercises to demonstrate how such data can be used to understand the connection between trademarks and firm characteristics and performance.

# 3 Data

This section describes the datasets that are used to match trademarks with firm-level data. It also provides an outline of the methodology for matching the trademarks to firms. Because the trademark data contains many separate files and a large number of variables, it is important to develop a strategy for using all the relevant information on trademarking by firms in the trademark files to facilitate the matching process.

## 3.1 Trademark data

The data on trademark activity comes from the USPTO Trademark Case Files Dataset (TCFD). This dataset was constructed by economists at the USPTO from case files made available by the USPTO on the `data.gov` website. The case files were organized and streamlined to form several electronic files that can be readily used by researchers to conduct large-scale analysis. The accompanying paper for data construction, Graham, Hancock, Marco, and Myers (2016), provides an excellent account of how the data was put together and a first look at some of the patterns of trademarking emerging from the data. The TCFD contains information on about 6.7 million USPTO trademarks applications and registrations for the period spanning 1870 to 2015. The information on trademarks includes, but is not limited to, ownership, assignment, prosecution history, classification, and renewal history.

In the United States, trademarks are subject to a use requirement, which requires that an entity use its mark on or in connection with the sales of goods or services in order to establish and maintain trademark rights. The use requirement derives from American common law and subsequent codification in federal statute.[7] An entity establishes and can enforce common law

---

[7]Under American common law, a trademark owner has the exclusive right to prevent unauthorized third parties from using the same or similar mark on goods and services where such use would likely cause confusion among

trademark rights solely by using a mark in commerce. A federal U.S. trademark registration confers considerable benefits beyond common law, specifically national-scope rights, prima facie evidence of ownership, access to the Federal court system to pursue injunctive relief and damages, and recordation with the U.S. Customs and Border Protection for monitoring the importation of infringing goods.

To file for a U.S. trademark registration, an applicant must specify the goods and services on which she uses or intends to use the trademark. The identified goods and services or "IDs" define the scope of trademark protection in the product space and generally cannot be overly broad or span multiple Nice classes.[8] Still, even within the same class, there can be considerable variation in the specificity of the goods and services listed and, thereby, the breadth of trademark protection.[9]

An applicant can file to register a trademark already in commercial use or for which she has a bona fide intent to use the trademark on the identified goods and services.[10] However, such so-called "intent-to-use applications" can only be registered after the owner deploys the trademark in commerce and provides a declaration and evidence supporting such use to USPTO.[11] A small, but growing, minority of applications are filed with the USPTO based on a foreign application or registration for the same trademark or an extension of an international registration to the U.S.[12]

---

consumers as to the source of the goods and services offered under the mark. An entity establishes trademark rights solely by using the mark on or in connection with goods and services. Registration at the state or federal level provides additional benefits but is not necessary for an entity to create and enforce common law trademark rights. The Lanham Act of 1946 ("Trademark Act") established the modern U.S. federal trademark registration system, providing for protection of trademarks used in commerce and registered with the USPTO. 15 U.S.C. §1051 et seq.

[8]The U.S. adopted the International Classification of Goods and Services under the Nice Agreement (the so-called "Nice Classification") in September 1, 1973. Prior to that date, the USPTO used a U.S. trademark classification system. Our match is restricted to trademarks registered under the Nice Classification system. See Trademark Manual of Examining Procedure (TMEP) §1401.

[9]See TMEP §1402.

[10]To file based on use in commerce under Trademark Act §1(a), the owner must submit a declaration stating that the mark is used in commerce that Congress can regulate, i.e., interstate commerce or commerce between the U.S. and foreign nations, as of the filing date. See TMEP §901.03. To file based on intent-to-use under §1(b), the applicant must have a bona fide intention to use the mark in commerce on the goods and services listed in the application in the near future. See TMEP §1101.

[11]Intent-to-use applications became available to applicants in November 1989 as a result of the Trademark Law Revision Act of 1988

[12]To file based on foreign application under §44(d), the applicant must own a foreign application that was filed within six months prior to the U.S. application for the same trademark and goods and services and claims a "priority" filing date based on the foreign application filing date. To file based on foreign registration under §44(e), the applicant must own a foreign registration of the same mark for the same goods and services from their non-U.S. country of origin. An application filed based on a foreign application cannot register in the U.S. until the

Authorized by international treaties, such applications can be registered prior to the trademark being used in U.S. commerce, however, generally only applicants with a foreign "country of origin" can obtain such U.S. trademark registration.[13] The owners with foreign addresses are excluded from the match with firm data.

Filing for a U.S. trademark registration costs a relatively small fee per class, and each application is subject to substantive examination by the USPTO for compliance with federal law and rules.[14] The USPTO determines if the applied-for trademark is legally protectable and that there is no "likelihood of confusion" with a previously registered trademark owned by another party.[15] If the examining attorney determines the applied-for mark is registrable, the USPTO publishes the trademark for a limited opposition period, during which third parties may file a formal opposition to the registration. Oppositions are fairly rare, instituted in less than 3% of published applications (Graham et al., 2013). Unopposed applications for trademarks already in use are issued a U.S. trademark registration. Allowed intent-to-use applications must first establish commercial use before the application can be registered.

A trademark owner can renew a U.S. trademark registration indefinitely so long as the trademark is used on the listed goods and services. The owner must provide proof of continued use and pay prescribe fees to the USPTO six years after registration and at each ten-year renewal event.[16] Failure to do so results in the registration being cancelled.

---

applicant establishes use in commerce under §1(a) or obtains a foreign registration under §44(e). To file based on an international registration under §66(a), the applicant must request extension of protection of an international registration to the U.S. through the International Bureau of the World Intellectual Property Organization.

[13]A U.S. applicant may file based on a foreign application but cannot obtain a U.S. registration unless the foreign application proceeds to registration and the applicant establishes that country as one of its countries of origin. An applicant's country of origin is "the country in which he has a bona fide and effective industrial or commercial establishment, or if he has not such an establishment, the country in which he is domiciled, or if he has not a domicile..., the country of which he is a national." See TMEP §1002.05.

[14]The application fee (as of the fee schedule effective January 14, 2017) is $600 per class for a paper application and $400 per class for a regular electronic application with additional discounts if the applicant submits other required forms electronically.

[15]An applied-for trademark can be refused as not registrable if, among other possible grounds, it is generic or merely descriptive, geographic, a surname, deceptive, inflammatory, a municipal, state, or national insignia, or the name, likeness, or signature of a living person used without their consent. 15 U.S.C. §1052. See TMEP §1200. Examining attorneys search existing registrations and pending applications for similar trademarks and assesses whether use of the applicant's trademark on the identified goods and services is likely to cause confusion among consumers. 15 U.S. C. §1052(d). See TMEP §1207.

[16]In the sixth year after the registration date, the trademark owner must maintain the registration by filing an affidavit or declaration of continued use, provide specimen(s) depicting use for the listed goods and services and pay prescribed fees.15 U.S.C. §§1058(a)(1). See TMEP §1604. Ten-year renewal terms were instituted for registrations issued on or after November 16, 1989. Registrations issued prior to that date had twenty-year terms until the first renewal event following that date. Thus, all live registrations are subject to ten-year terms as of

The coverage of the TCFD varies over the years. Graham, Hancock, Marco, and Myers (2016) indicate that there is little coverage of classification, prosecution event history, and owner records for trademark registrations issued before 1962.[17] Only about 16% of the registered trademarks before 1962 have such data items populated. The coverage improves for registrations issued during the period 1962-1977, to about 72-81% depending on the specific data item. Substantial improvement in coverage for filings and registrations occurs after 1977, and coverage becomes nearly 100% after 1982. Figure 1 gives a time series of the number of trademark filings and registrations in our sample period. The key data items are populated at a rate of 89% or more for the period 1977-2015. Fortunately, as discussed below, firm level longitudinal micro data is available in the U.S. Census Bureau starting in 1976. Therefore, much of the well-covered trademark activity over the period 1977-2015 in the TCFD coincides with the coverage of the longitudinal micro data on firm characteristics for that same period. However, the left censoring of the firm data at 1976 and the relatively lower coverage of the trademark data before 1977 implies that some firms born before 1976 may not match with any trademark activity, even though they may have filed for or registered a trademark before 1976. Therefore, the focus of the analysis will be on firms that were born in 1976 and following years.

## 3.2   Firm data

The trademark data is matched with the U.S. Census Bureau's Business Register (BR), which is the universe of administrative data on businesses in the United States. The BR is also the sampling frame for the Census Bureau's economic surveys. It contains information on a firm's employment, payroll, and revenues, as well as geography and industry classification. The analysis requires longitudinal data to track firms over time and to understand when in the life-cycle of a firm trademark activity occurs. The Longitudinal Business Database (LBD) provides a longitudinally linked version of BR at the establishment level for the period 1976-2014. The LBD allows identification of entry and exit of firms based on the first and last observations of a firm identifier in administrative data on payroll or employment. Since the LBD coverage starts in 1976, there is no age information for establishments of a firm in 1976. The analysis is therefore restricted to firms born on or after 1977.

The primary measures of firm size used in this paper are employment and revenue. Previous work has linked observations in BR on firm revenue longitudinally to construct a longitudinal

---

November 16, 2009. Registrants must pay separate maintenance and renewal fees for each class in the registration. The per class fee (as of the fee schedule effective January 14, 2017) is $125 for electronic filing ($225 for paper filing) an affidavit or declaration of use and $300 for electronic filing ($500 for paper filing) a renewal application. 15 U.S.C. §1059(a). See TMEP §1606.

[17]See Table 1 in Graham, Hancock, Marco, and Myers (2016).

firm revenue database. Details on the construction of this dataset is in Haltiwanger, Jarmin, Kulick, and Miranda (2016). The analysis uses this dataset to examine the connection between firm revenue growth and various trademark activities.

In addition to the BR, the matching process utilizes the Integrated Longitudinal Business Database (ILBD). Even after matching the trademark data to the BR, there are many trademarks left unmatched with a firm in the bR. Many of these trademarks appear to be owned by individuals, as indicated by the person names in their records in the TCFD. The ILBD provides the longitudinally link observations for individuals' income tax records pertaining to the Schedule C filings that indicate earnings for a business.[18] The ILBD non-employer universe is used to match person names across the unmatched trademarks and tax records to the extent possible.

## 3.3   Matching process

This section describes the matching of records in the TCFD to the BR. Because there are no unique identifiers, such as an Employer Identification Number (EIN), that are shared between the TCFD and the BR, the two datasets have to be brought together using name and address matching techniques. The matching strategy for the trademark data from TCFD is to extract all unique combinations of name and address information from the numerous places in the TCFD where they are observed. Name and address information are collected at different times during a trademark's lifecycle often for the same business entity, which provides additional variation in the strings to be matched. Not only are there multiple instances of the name and address information for one or more businesses associated with a trademark, but also there are different types of business names. For instance, one observes "former", "doing business as", and "composed of "qualifiers for names. To keep the set of names and addresses to be matched with BR as large as possible, a unique set of name and address combinations within each TCFD record across all name types is generated.

To prepare the trademark and BR data for matching, some cleaning of the name and address strings is done. Common strings that carry little identifying information, and which may not be present in both datasets, such as "formerly known as", are removed from both datasets. To improve the quality of the matches to the BR, it is also important to identify the year(s) in which a name and address combination is observed in the TCFD. For this purpose, the minimum and maximum dates associated with each name and address combination are extracted. These dates include the filing, registration, and publication dates for each TCFD record.

Matching to the BR is done in several steps. First, an initial subset of potential matches are identified based on a relatively loose name-only match criteria. The matches not meeting this

---

[18]See Haltiwanger and Jarmin (2007) for details on the construction of the ILBD.

minimum criteria are not considered. This initial match is then strengthened using conditions utilizing both name and address fields. These conditions include nine different match categories, based on whether the match is on name in conjunction with zip code, state or city only, or any combination of them, and finally, name only. A string comparator is then used to disambiguate and improve match quality. Additionally, the minimum and maximum trademarking years, and the minimum and maximum years each firm is observed in the BR, are used to filter matches. Finally, cases that are not matched to the BR are matched to the non-employer business register (ILBD).[19] Ultimately, over 70% of trademarks and owner records are matched to at least one firm. Roughly 55% of these matches are to a single firm. The analysis to follow will use only unique matches. The TCFD-BR match rates are stable across several trademark characteristics.
[20]

Once names and addresses from the TCFD are matched to the BR, a firm-level panel for trademarking firms is constructed using the LBD. Establishment characteristics such as employment and payroll are aggregated to the firm-level using the longitudinally consistent firm identifiers available in the LBD. The TCFD-BR bridge is then used in along with trademark attributes found in the TCFD to characterize each matched firm's trademarking activity. This includes the number of trademarks each firm filed, registered, and published each year.

## 3.4   Critical events and actions

As a prerequisite to any analysis of trademarks, it is important to identify critical events pertaining to trademarking over the life-cycle of firms. While there are many key events in a trademark's life-cycle, the two events are particularly important in understanding the nature of selection into the pool of firms that possess trademarks. The first critical event this paper focuses on is the first ever filing for a trademark by a firm. The filing date of a trademark is available in the TFCD. In principle, one can use the filing date variable for each trademark to identify the timing of the first ever application for a trademark. However, the process is a bit more complicated, because a trademark can have several applications over time corresponding to different classes, and firms may also have several trademarks. For instance, a firm may initially apply for a trademark for computer services, but may later extend this trademark to another class, such as customized computer software. For cases where there is more than one application for a trademark or multiple trademarks, the minimum filing date across all filing dates is used to identify the first ever filing for a trademark. Note that a trademark filing can predate a firm's entry date in the Longitudinal

---

[19]At this stage non-employer matches are used primarily to better understand the quality of matches to the employer universe. Future work will investigate matches to non-employer businesses and their growth and transition dynamics.

[20]For additional detail on the data construction and matching see Appendix.

Business Database, as some firms may apply for a trademark even before they become employers (pay their first wage or hire their first employee), or register as a non-employer business.

Another critical event studied here is the registration of a trademark. A large number of trademark applications are abandoned even before reaching registration. It is important to note that in most cases firms have to demonstrate commercial use of a trademark before a registration is granted by the USPTO. Therefore, registration contains valuable information about commercial activity associated with a trademark application. In other words, registration is a *necessary* condition for commercial use of a trademark. The event of registration represents another layer of selection for firms which have filed a trademark application. One expects the firms that file and eventually register to differ from those that file but abandon the application before registration. Again, identifying the first registration date for a trademark with multiple registrations over time is difficult, since there may be several registrations for different classes under the same trademark. As in the case of filing, the earliest date associated with all registrations of a trademark is used to identify the timing of the first registration for a trademark.

# 4   Analysis

This section presents some preliminary findings on the nature of trademarking firms using the matched data. The main goal is to understand how trademarking firms differ from others, and whether trademarking is correlated with some firm performance measures, such as growth in employment or revenues. The section also presents some basic correlations between trademarking, patenting and R&D activity at the firm level. The analysis is highly descriptive in nature and does not seek to establish any causality between trademarks and firm performance and innovation, a pursuit left for future work.

## 4.1   What kind of firms trademark?

This section examines the patterns of trademarking by firm size and age, the two key state variables for firms that are frequently used in the literature to correlate with firm outcomes and performance. For the rest of the analysis, only the firms born on or after 1976 are considered to ensure that the left censoring of the firms born before 1976 does not affect the results.

### 4.1.1   Trademarking by firm size and age

Consider, first, some baseline statistics on firm counts by firm types in the matched data, as shown in Table 1. Four firm types are identified for each year: firms with a trademark (applied for a trademark in a prior year), firms that trademark (applied for a trademark in the current

year), firms that trademark the first time (firms that apply for a trademark the first time in the current period) and non-trademarking firms (firms that have never applied for a trademark as of the current year). The column labelled "Matched TMs" in Table 1 shows the total trademark counts in each year starting in 1977. The growth in trademarks is significant. The trademark count in a year goes from roughly 5,800 in 1977 to more than 65,000 in 2013. At the same time, relatively few firms hold a trademark. The percentage of firms that hold a trademark in any given year increases from 0.5% in 1977 to nearly 3% in 2013. The number of firms which apply for a trademark at any given year also grows steadily from 0.1% in 1977 to 0.5%. Finally, the average number of trademarks per firm in a year increases from 1.7 to 2.1 between 1977 and 2013.

Table 2 reports the average size (employment) and age for firms by trademark activity. Note that age is left-censored. A first observation based on Table 2 is the relatively large size of firms that hold trademarks. For instance, firms that have at least one trademark in 1977 ("Firms with a TM") are nearly thirty times larger than non-trademarking firms ("Firms with no TM"), and firms that apply for trademarks in 1977 ("Firms that TM") are almost forty times larger on average. The average size of trademarking firms declines over the sample period substantially. The average size of firms that apply for trademarks for the first time is around 9 employees over the sample period, but declines to some extent between 1977 and 2013. Firms with a trademark are also significantly older than firms with no trademark on average. First-time applicants for a trademark typically do so around their birth year, as indicated by their average age of less than one year.

Figure 2 contains the cumulative distribution function of firm size (employment) at the time of a firm's first trademark filing. Most first-time trademarkers are small (nearly 80%), with less than 20 employees. However, there is a long tail to this distribution, implying that there are also many larger firms that trademark the first time.

Figure 3 plots a local polynomial smoothing for the likelihood of having made a trademark application for firms in various employment size bins for the year 2013. The figure shows a gradually upward sloping curve that begins to get steeper when the firm size is greater than about 400 employees. The figure shows that when firms are sufficiently large, the vast majority of them have applied for at least one trademark over their life-cycles.

First-time trademarking is also concentrated in young firms, as Figure 4 indicates. Most firms (about 75%) are less than 5 years old when they file for their first trademark, indicating that many firms file for a trademark earlier in their life-cycle. As in the case of firm size, there is a large mass of firms beyond age five, indicating the presence of many old firms that apply for a trademark the first time.

Based again on the 2013 cross-section of firms, the local polynomial estimate in Figure 5

shows that as firms get older, they are more likely to apply for a trademark, with the likelihood of trademarking remaining essentially unchanged once the firm reaches about 12 years of age.

Figure 6 provides a time series of filings for a trademark by firm age. Note the spike that occurs during the dot-com boom (1999-2000) when the number of filings by startups (firms with age zero) increases sharply. The figure shows that the number of filings by firms aged 5 and younger are significantly larger than the number of filings by older firms. It also indicates that firms aged 0-5 tend to constitute a larger fraction of first-time trademarkers over time.

### 4.1.2 Trademarking by industry

The intensity of trademarking activity depends on the industry and sector of the firm. The distribution of trademarks across various industry groups is shown in Figure 7. Firms are classified into industries based on the industry where their largest share of employment falls in. The largest share of trademarks is accounted by firms in manufacturing and wholesale, followed by professional services, retail, and information.

Table 3 presents a 2013 snapshot of the average size and age of firms by trademarking activity in various sectors. There is considerable variation across industry groups. For example, firms in wholesale, professional services and other services apply for trademarks at a significantly smaller size than firms in other industries like agriculture and education. Firms in retail and entertainment apply for trademarks for the first time much earlier in their life-cycle than other industries. Part of these differences are clearly driven by the differences in size and age distributions of firms across sectors.

Figure 8 shows the distribution of all firms that have a trademark by broad industry groups over time. Starting with the 1980's, there is an expansion of trademarking by professional services firms. In the 1990's, trademarking by retail and wholesale firms also start to increase. It is notable that the "Other Firms" category, which includes firms in all other industries, has also grown substantially since the 1980's, while trademarking by manufacturing firms remains relatively stable. The dot-com era spike in trademarking activity is clearly visible right before 2000, after which the growth in many categories has been relatively stable.

The descriptive analysis of the matched data so far indicates that very few firms seek out trademarks. Fewer than 3% of all firms have applied for a trademark in their life-cycles, and less than 0.5% of firms apply for a trademark in any given year. Firms that have applied for trademarks are significantly older and larger than firms that have not. First-time trademarkers typically apply for their trademark very early in their life-cycle. Nearly 80% of firms that apply for a trademark for the first time do so when they have fewer than 20 employees. There is also significant variation in the size and age of trademarking firms across industries. The next subsection explores some of the determinants of trademarking.

14

### 4.1.3 Some determinants of trademarking

We consider next a logit specification for the first time filing and registration, where the dependent variables include size and age categories and industry-year fixed effects. The results of the logit analysis is in Table 4. The estimated coefficients associated with various size and age categories indicate that the likelihood of a firm filing or registering a trademark monotonically declines with age (controlling for size), and monotonically declines with size (controlling for age).

To explore the connection between trademark activity and firm size further, consider next the following regression

$$
\begin{aligned}
SIZE_{it} \;=\; & \alpha + \beta_1 FILE_i * PRE\_FILE_{it} + \beta_2 FILE_i * POST\_FILE_{it} \qquad (1) \\
& + \gamma_1 REGISTER_i * PRE\_REGISTER_{it} + \gamma_2 REGISTER_i * POST\_REGISTER_{it} \\
& + \lambda AGE_{it} + INDUSTRY_i + YEAR_t + \varepsilon_{it},
\end{aligned}
$$

where $SIZE_{it}$ is a measure of firm size (employment or revenue), $AGE_{it}$ is firm age, $INDUSTRY_i$ and $YEAR_t$ are industry and year fixed effects, and $\varepsilon_{it}$ is an error term. The other variables capture the effects of trademark filing and registration on size. $FILE_i$ is an indicator that takes on a value of one if a firm has ever filed for a trademark. Similarly, $REGISTER_i$ is an indicator that equals one if a firm has ever completed a registration of a trademark. These two variables are aimed at measuring the size premium associated with filing and registration activity. In addition, $PRE\_FILE_{it}$ is an indicator that is turned on only for the years for which a firm has not yet filed for a trademark, conditional on filing a trademark at some point in time ($FILE_i = 1$). For the years after filing, the indicator variable $POST\_FILE_{it} = 1 - PRE\_FILE_{it}$ captures any size premium for filers after the year of filing. $PRE\_REGISTER_{it}$ and $POST\_REGISTER_{it}$ are defined analogously. The omitted category in the regression is the set of firms that never file for a trademark, and hence, never register one.

Note that registration is a necessary condition for commercial use of a trademark. Compared to mere filing for a trademark, the event of registration can be more strongly correlated with a firm's intentions to use the trademark for commercial purposes. The coefficients $\beta_k$ and $\gamma_k$ ($k = 1, 2$) capture how firms that file for trademarks and register them differ in size from others, controlling for age, industry, and year.

The results from the regression in (1) is in Table 5. The coefficient estimates for pre- and post-filing periods, as well as pre- and post-registration periods indicate that firms that file or register for a trademark tend to be larger before and after these events compared to the firms that do not file for a trademark over the sample period. It is interesting to note that there is an additional size premium for registrants conditional on filing. However, the regression in (1) is exploratory in nature, and does not seek to identify any causal effects of trademarking activity on firm size.

### 4.1.4 The number of trademarks per firm

One measure of how intensely firms trademark is the number of trademarks held by a firm, conditional on having a trademark. Figure 9 shows the kernel density estimate of the counts of trademarks registered by firms. Most firms that have applied for a trademark have done so for a small number of trademarks (less than 3), but there are also many firms with a large number of trademarks out in the right tail of the distribution. It is not uncommon for firms to expand trademark registrations into related and unrelated business lines to take advantage of the reputation and goodwill built earlier.

The pattern in Figure 9 has been very stable over the years, with almost no change in the distribution of trademark holdings over time. While Table 1 indicates that the average number of trademarks applied for by firm has increased from 1.6 to 2.1 over a 35-year time horizon, this shift is mainly caused by firms at the right tail of the distribution . Across all years, 75% of firms that have applied for a trademark have done so for less than two trademarks.

## 4.2 Trademarking and firm growth

This section presents some preliminary findings on the connection between firm growth and trademark activity. Figure 10 plots the pre-trademark performance with the post-trademark performance for a balanced panel of firms that trademark using a local polynomial smoothing in the pre- and post-periods.[21] The estimated polynomial measures the within-firm employment and revenue trends, normalizing the employment and revenue levels at the time of trademark filing ($t = 0$) to 100. In the run-up to the trademark filing year, the mean employment is mostly flat and slowly increasing. Once the firm files for a trademark however, the employment level steadily increases by approximately 20% in the two-years following the filing. Revenue, which is generally more volatile than employment, stays mostly flat in the pre-period, before increasing and following a similar trend as employment in the post-period.

Figure 11 plots the employment and revenue growth before and after trademark filing by size, sector, and age. The top row suggests that the growth from trademarking occurs mostly in smaller firms, with firms that have fewer than 10 employees and fewer than $1M in revenue two years before trademark filing growing the fastest in the post-period. Across sectors, there is little variation, but it appears that professional service firms benefit the most in terms of growth following the trademark. Finally, young firms (defined as ages 2-5 years old at the time of filing) also grow the fastest, with middle and older firms staying mostly flat in the post-period.

It is important to note that the event study described above uses the firms that trademark

---

[21] The sample is a balanced panel, conditional on survival 2 periods before/after initial trademark filing. The plot specifies a Gaussian 3rd degree polynomial smooth function over the entire period with 95% confidence bands.

at age two and later. The majority of firms trademark at an earlier age, but it is not possible to construct an event study for such firms as they do not have a sufficiently long period of activity before trademark filing. Instead, Figure 12 which shows the growth rates of firms who apply for a trademark in the first year of existence and compares them with the cohorts born in the same year that never trademark. The figure suggests that firms which trademark in their first year of existence tend to have a different employment profile over their life-cycles compared to those firms that do not trademark at all. Trademarking firms tend to be larger even in their first year compared to the firms that do not trademark in their first year. There is also a steeper employment-age profile for firms that file for a trademark in their first year. These patterns suggest that trademarking entrants tend to differ from other entrants, hinting at a strong selection pattern or trademarking firms even at birth.

To understand the connection between firm growth and trademark activity further, the following regression is also considered

$$
\begin{aligned}
GROWTH_{it+1} \;=\; & \alpha + \beta_1 FILE_i * PRE\_FILE_{it} + \beta_2 FILE_i * POST\_FILE_{it} \\
& + \gamma_1 REGISTER_i * PRE\_REGISTER_{it} + \gamma_2 REGISTER_i * POST\_REGISTER_{it} \\
& + \lambda AGE_{it} + \phi SIZE_{it} + INDUSTRY_i + YEAR_t + \varepsilon_{it},
\end{aligned}
\tag{2}
$$

where

$$
GROWTH_{it+1} = 2 \times \left( \frac{SIZE_{it+1} - SIZE_{it}}{SIZE_{it+1} + SIZE_{it}} \right)
$$

is the growth rate based on a size measure $SIZE_{it}$, which is either employment or revenue. The growth rate $GROWTH_{it+1}$ can be calculated for entering and exiting firms, for which $SIZE_{it} = 0$ and $SIZE_{it+1} = 0$, respectively.[22] The other variables are defined as in regression (1).

Table 6 presents the results from the regression in (2). The coefficient estimates indicate that firms that file for a trademark and those that register have higher employment and revenue growth rates before and after the filing and registration events, compared with the firms that do not file at all over the sample period. Again, these regressions are descriptive in nature and do not indicate any causality between trademarking activity and firm growth.

## 4.3 Trademarking and innovation

The incidence of a new trademark application by a firm can be associated with a variety of firm activities. One of these activities may be the introduction of a new product or service, which can be classified as an innovation. To provide a preliminary exploration of the link between trademarks and innovation, Table 7 provides a matrix on the incidence of patents and R&D

---

[22]See Davis, Haltiwanger and Schuh (1996) for this growth measure.

activity for trademarking firms, as well as the incidence of trademarking within firms with patents and R&D. The data for patenting and R&D come from the recently constructed U.S. Census Bureau longitudinal business patent database to identify patents assigned to employer businesses, as well as two Census R&D surveys: the Business Research and Development and Innovation Survey (BRDIS) and the Survey of Industrial Research and Development (SIRD). All firms that own patents and have ever conducted R&D are identified and merged with the trademark-firm matched data. Table 7 provides various correlations across the three types of activity (trademarking, patenting and R&D) for the period of 1998-2008.

Table 7 indicates that 7.6% of firms that have applied for a trademark also have a patent, and 3.8% of trademarking firms also conduct R&D between 1998 and 2008. While these rates may seem low, they are significantly higher than the economy-wide rates of patenting and R&D by all firms, especially considering that many trademarking firms are associated with sectors which are typically not innovation-intensive, such as retail and wholesale. Turning to patenting firms, the incidences of trademarking and R&D are significantly higher. 16.5% of patenting firms also have a trademark, and 22.5% of patenting firms also conduct R&D. Finally, for R&D performing firms, the matrix in Table 7 indicates that nearly 20% have applied for a trademark and a little over half own a patent. Overall, the high incidence of trademarking by patenting and R&D conducting firms may indicate that firms that engage in innovative activities may seek to protect their investment into new products and services by establishing brand names via trademarks. This hypothesis needs further testing. One task for future work is to explore the exact timing of patent and trademark applications at the firm level, which can shed light on whether trademarks tend to follow successful innovative activity at the firm level.

# 5   Conclusion

Research on trademarking activity by firms in the U.S. has lagged behind that in other countries due to a lack of comprehensive firm-trademark matched data. This paper has reported on the construction of a firm-level database that allows tracking of activity related to trademarks over the life-cycle of a firm. The dataset brings together the USPTO's Trademark Case Files Dataset and the US Census Bureau's Business Register for the period 1976-2014. Using the dataset, it is possible to analyze when a firm first applies for a trademark and when it registers a trademark, which requires demonstration of commercial use of the trademark. These key activities are tied to firm characteristics such as firm size and age to understand which type of firms trademark.

The longitudinal aspect of the firm data allows for an exploration of how firm growth is related to trademark activity. The analysis indicates that firm employment and revenue growth rates are higher for firms that have filed for a trademark, and even higher for those firms that

have completed registration of a trademark. These findings, however, do not point necessarily to a causal role for trademarks in firm performance. It is important to look for ways to identify in future work whether trademarks indeed have a separate effect on firm performance.

The dataset opens up several possibilities for future research. In particular, the incidence and timing of various stages of trademark activity in the data can be used to test theories of goodwill building and reputation, as well as theories of marketing. For instance, if trademarks are indeed able to protect goodwill of firms, one would expect to observe that firms with trademarks invest more heavily in advertising and product portfolio expansion. Another avenue for future research is to understand in more detail the link between firm-level innovative activity and trademarking. Whether trademarks are lagging or leading indicators of innovation is an open question. Answers can be provided using the dataset constructed here in conjunction with other datasets that contain information of firm level innovative activity, such as patenting, R&D, and product introductions.

# References

[1] Baroncelli, Eugenia, Carsten Fink, and Beata S, Javorcik. 2005. "The Global Distribution of Trademarks: Some Stylized Facts." *The World Economy*, 28: 765-782.

[2] Davis, Steven, Haltiwanger, John, and Scott Schuh. 1996. *Job Creation and Destruction*. Cambridge, MA: The MIT Press.

[3] Economides, Nicolas. 1988. "The Economics of Trademarks.", The Trademark Reporter, 78: 523-539.

[4] Fazio, Catherine, Jorge Guzman, Fiona E. Murray, and Scott Stern. 2016. "A New View of the Skew: A Quantitative Assessment of the Quality of American Entrepreneurship" In *Kauffman Foundation New Entrepreneurial Growth*, Kansas City, MO: February 2016.

[5] Foster, Lucia, Grim, Cheryl, and Nikolas Zolas. 2016. "A Portrait of Firms that Invest in R&D." US Census Bureau Center for Economic Studies Paper No. CES-WP-16-41

[6] Graham, Stuart, Galem Hancock, Alan Marco, and Amanda F. Myers. 2016. The USPTO Trademark Case Files Dataset: Descriptions, Lessons, and Insights.

[7] Greenhalgh, C., M. Longland, and D. Bosworth. 2003. "Trends and Distribution of Intellectual Property: U.K. and European Patents and U.K. Trade and Service Marks, 1986-2000." *Report for the U.K. Intellectual Property Office*.

[8] Greenhalgh, C., and M. Rogers. 2008. "Intellectual Property Activity by Service Sector and Manufacturing Firms in the U.K., 1996-2000." *The Evolution of Business Knowledge,* Oxford University Press.

[9] Haltiwanger, John, and Ron Jarmin. 2007. "Integrated Longitudinal Business Database: Data Overview." 2007 Kauffman Symposium on Entrepreneurship and Innovation Data

[10] Haltiwanger, John, Jarmin, Ron, Kulick, Robert, and Javier Miranda. 2016. "High Growth Young Firms: Contributions to Job, Output, and Productivity Growth", forthcoming in *Measuring Entreprenuial Businesses: Current Knowledge and Challenges*, NBER Book Series Studies in Income and Wealth.

[11] Jensen, P. H. and E. Webster. 2004. "Patterns of Trademarking Activity in Australia." *Australian Intellectual Property Journal*, 15:112-126.

[12] Landes, William and Richard Posner. 1987. "Trademark Law: An Economic Perspective," *Journal of Law and Economics*, 30: 265-309.

[13] Landes, William and Richard Posner. 1988. "The Economics of Trademark Law," *The Trademark Reporter*, 78: 267-306.

[14] Sandner, Philipp. 2009. "The Identification of Trademark Filing Strategies: Creating, Hedging, Modernizing, and Extending Brands." *The Trademark Reporter*, 99: 1257-1298.

[15] Sandner, Phillipp and Joern Block. 2011. "The Market Value of R&D, Patents, and Trademarks." *Research Policy*, 40: 969-985.

[16] Zolas, Nikolas, Lybbert, Travis, and Prantik Battacharyya. 2016. "An 'Algorithmic Links with Probabilities' Concordance for Trademarks with an Application Towards Bilateral IP Flows", forthcoming in *The World Economy*, doi: 10.1111/twec.12382.

Table 1. Trademark and firm counts

| Year | Matched TMs | All firms | Firms with a TM | Firms that TM | Firms that TM the first time | Firms with no TM | TMs per firm |
|------|------|------|------|------|------|------|------|
| 1977 | 5,760 | 3,653,800 | 16,770 | 3,380 | 3,380 | 3,637,000 | 1.71 |
| 1982 | 10,990 | 3,785,500 | 26,280 | 6,290 | 4,580 | 3,759,200 | 1.75 |
| 1987 | 11,920 | 4,418,100 | 36,770 | 7,250 | 4,780 | 4,381,300 | 1.65 |
| 1992 | 19,610 | 4,640,400 | 51,400 | 10,550 | 6,200 | 4,589,000 | 1.86 |
| 1997 | 39,340 | 5,052,900 | 74,740 | 18,160 | 10,360 | 4,978,200 | 2.17 |
| 2002 | 42,090 | 5,258,600 | 100,630 | 19,710 | 10,350 | 5,157,900 | 2.14 |
| 2007 | 60,020 | 5,638,200 | 128,000 | 27,170 | 14,600 | 5,510,200 | 2.21 |
| 2013 | 65,170 | 5,722,400 | 161,930 | 30,480 | 15,400 | 5,560,500 | 2.14 |

Notes: All figures have been rounded for disclosure purposes.

Table 2. Average employee size and age of firms by trademark activity

| | Firms with a TM | | Firms that TM | | Firms that TM the first time | | Firms with no TM | |
|---|---|---|---|---|---|---|---|---|
| Year | Size | Age | Size | Age | Size | Age | Size | Age |
| 1977 | 364.9 | 1 | 504.4 | 0.8 | 12.2 | 0.8 | 12.9 | 0.8 |
| 1982 | 273.8 | 5 | 330.2 | 3.8 | 9.5 | 0.1 | 13.9 | 3.8 |
| 1987 | 225.2 | 8 | 260.3 | 5.3 | 10.8 | 0.1 | 14.1 | 5.6 |
| 1992 | 182.3 | 10.2 | 243.8 | 6.9 | 9.8 | 0.1 | 14.3 | 7.3 |
| 1997 | 167.9 | 11.5 | 251 | 7.7 | 9.6 | 0.1 | 14.4 | 8.4 |
| 2002 | 146.1 | 12.9 | 227.2 | 8.9 | 9.3 | 0.2 | 14.2 | 9.5 |
| 2007 | 132.8 | 14.3 | 208.8 | 9.7 | 8.7 | 0.2 | 14 | 10.3 |
| 2013 | 114.8 | 16.2 | 193.7 | 10.8 | 4.6 | 0.1 | 12.9 | 11.6 |

Notes: All figures have been winsorized (at 99.9% level) for disclosure purposes.

Table 3. Average employee size and age of firms by trademark activity - sectors, 2013

| Sector | Firms with a TM | | Firms that TM | | Firms that TM the first time | | Firms with no TM | |
|---|---|---|---|---|---|---|---|---|
| | Size | Age | Size | Age | Size | Age | Size | Age |
| Agriculture, Mining and Utilities | 656.7 | 19.3 | 1308.1 | 13.4 | 5.6 | 0.3 | 16.7 | 12.8 |
| Constr., Transp., and Warehousing | 160.5 | 17.4 | 313.6 | 11.3 | 2.8 | 0.1 | 8.6 | 11 |
| Manufacturing | 137.9 | 21.6 | 246.1 | 15.2 | 13.2 | 0.5 | 29.9 | 15.2 |
| Wholesale | 52.1 | 16.8 | 73.9 | 12.1 | 4.1 | 0.4 | 12.6 | 13.2 |
| Retail | 139.7 | 14.1 | 253.3 | 8 | 3.4 | 0.1 | 10.4 | 11.4 |
| Information | 92 | 14.2 | 207.7 | 9.8 | 9.9 | 0.3 | 17.4 | 10.4 |
| Finance and Real Estate | 190.8 | 16 | 413.2 | 11.9 | 2.9 | 0.1 | 7.6 | 11.5 |
| Professional Services | 41.3 | 12.3 | 59.3 | 7.4 | 2.9 | 0.2 | 7.1 | 10.1 |
| Management and Admin. Support | 196.6 | 15.4 | 334.8 | 11.3 | 6.7 | 0.1 | 19.8 | 9.6 |
| Education and Healthcare | 261.4 | 16.2 | 376.2 | 11.6 | 6.2 | 0.1 | 22.1 | 12.2 |
| Entertainment, Accomm., and Food | 125.4 | 14.4 | 204.7 | 8.3 | 7.4 | 0.1 | 18.1 | 8.9 |
| Other Services | 42.6 | 19.9 | 73.2 | 14.8 | 2.4 | 0.1 | 7 | 15.7 |

Notes: All figures have been winsorized (at 99.9% level) for disclosure purposes.

Table 4. Logit analysis of first filing
and registration for a trademark

| | First time: | |
| | Filing | Registration |
| --- | --- | --- |
| 0-5 emp | −2.870*** | −3.074*** |
| | [0.013] | [0.013] |
| 5-9 emp. | −2.336*** | −2.470*** |
| | [0.014] | [0.014] |
| 10-19 emp. | −2.020*** | −2.100*** |
| | [0.014] | [0.014] |
| 20-49 emp. | −1.661*** | −1.685*** |
| | [0.014] | [0.014] |
| 50-99 emp. | −1.362*** | −1.346*** |
| | [0.015] | [0.016] |
| 0-2 years | 1.883*** | 1.705*** |
| | [0.010] | [0.011] |
| 3-4 years | 1.171*** | 1.294*** |
| | [0.011] | [0.012] |
| 5-9 years | 0.848*** | 0.904*** |
| | [0.011] | [0.011] |
| 9-14 years | 0.506*** | 0.519*** |
| | [0.012] | [0.013] |
| 15-19 years | 0.281*** | 0.294*** |
| | [0.013] | [0.016] |
| | | |
| $N$ | $> 100M$ | $> 100M$ |

Notes: All regressions include industry-year fixed effects.
Standard errors in prantheses. *,**,*** indicate
significance 1,5, and 10% repectively. Omitted
categories are 500+ employees and 25+ years.
Coefficients  for other categories are not shown
to avoid disclosure (size categories 100-249,
250-499; age categories 20-25, 25+).

Table 5. Regression analysis of the relation
between firm size and trademark activity

| | Firm Size Measure | |
| --- | --- | --- |
| | ln(Employment) | ln(Revenue) |
| $\beta_1$ | 0.617*** | 0.846*** |
| | [0.014] | [0.012] |
| $\beta_2$ | 0.593*** | 0.805*** |
| | [0.011] | [0.009] |
| $\gamma_1$ | 0.314*** | 0.370*** |
| | [0.012] | [0.011] |
| $\gamma_2$ | 0.318*** | 0.339*** |
| | [0.011] | [0.010] |
| $\lambda$ | 0.189*** | 0.196*** |
| | [0.009] | [0.010] |
| | | |
| $N$ | $> 100M$ | $> 100M$ |

Notes: All regressions industry and year fixed effects.

Standard errors cluatered by industry in prantheses.

 *,**,*** indicate significance 1,5, and 10% repectively.

Table 6. Regression analysis of the relation
between firm growth and trademark activity

| | Firm Growth Measure | |
| | Employment Growth | Revenue Growth |
| --- | --- | --- |
| $\beta_1$ | 0.093*** | 0.099*** |
| | [0.002] | [0.002] |
| $\beta_2$ | 0.053*** | 0.023*** |
| | [0.002] | [0.001] |
| $\gamma_1$ | 0.061*** | 0.068*** |
| | [0.002] | [0.002] |
| $\gamma_2$ | 0.006*** | 0.013*** |
| | [0.001] | [0.001] |
| $\lambda$ | $-0.053$*** | $-0.021$*** |
| | [0.000] | [0.001] |
| $\phi$ | $-0.357$*** | $-0.010$*** |
| | [0.000] | [0.001] |
| | | |
| $N$ | $> 100M$ | $> 100M$ |

Notes: All regressions include industry and year fixed effects.

Standard errors clustered by industry in parantheses.

*,**,*** indicate significance 1,5, and 10% repectively.

Table 7. Trademark and innovation incidence matrix

| Probability a firm has ⟶ | Trademark | Patent | R&D |
|---|---|---|---|
| Firms with a trademark | 100 | 7.6 | 3.8 |
| Firms with a patent | 16.5 | 100 | 22.5 |
| Firms with R&D activity | 19.7 | 53.3 | 100 |
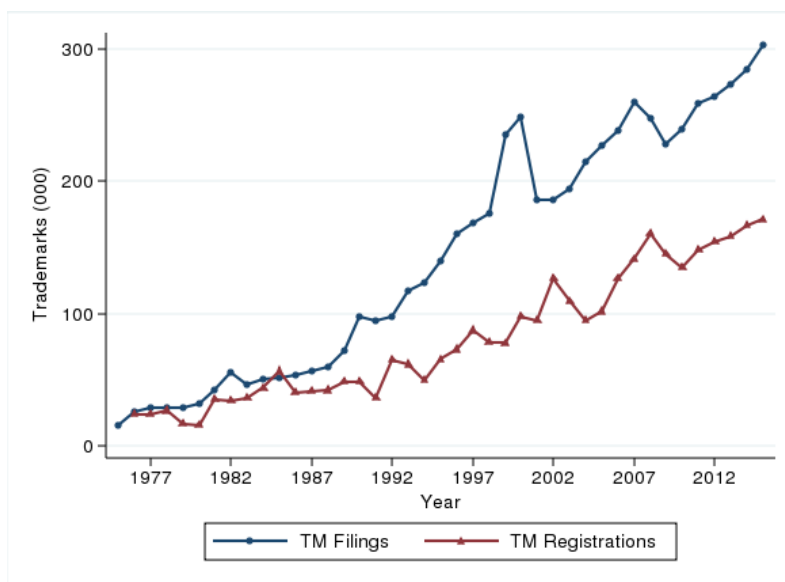
Figure 1: The number of trademarks over time
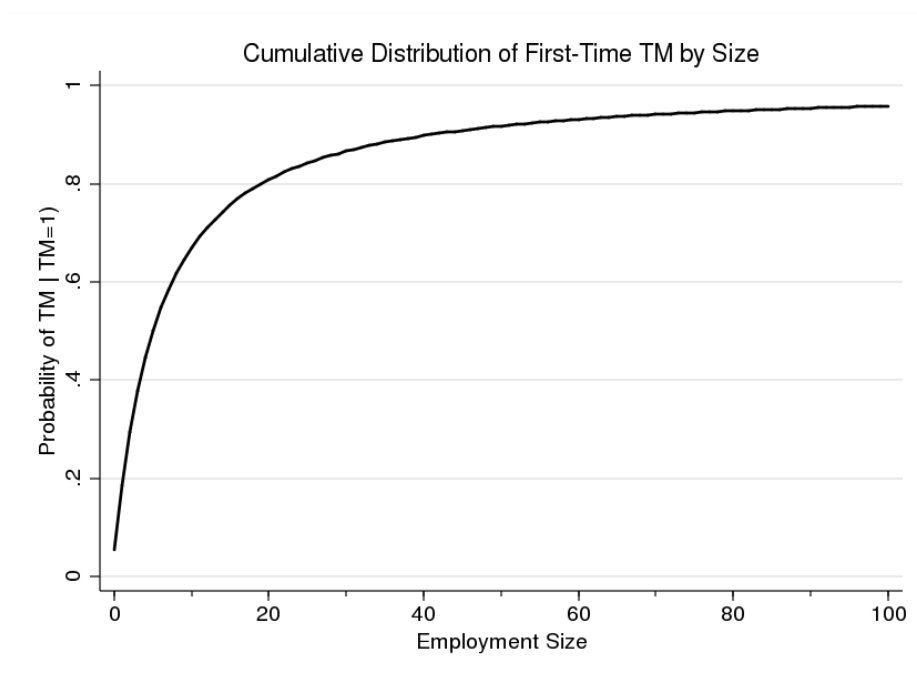


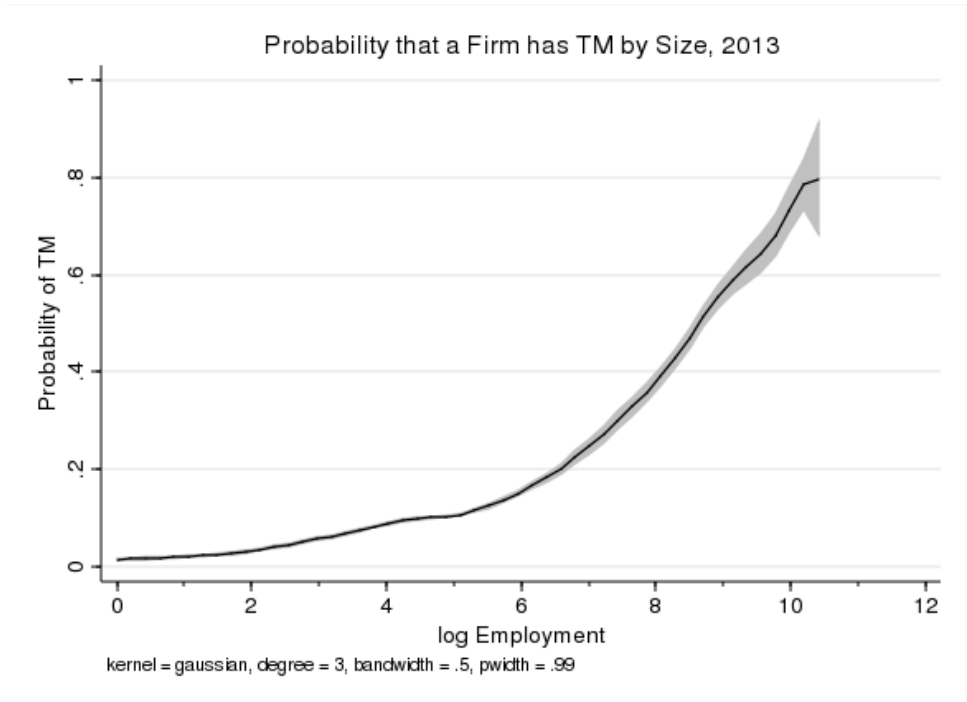Figure 2: Cumulative distribution function of firm size (employment) for first-time trademarkers

Figure 3: The relationship between the likelihood of having applied for a trademark and firm size - local polynomial smoothing.
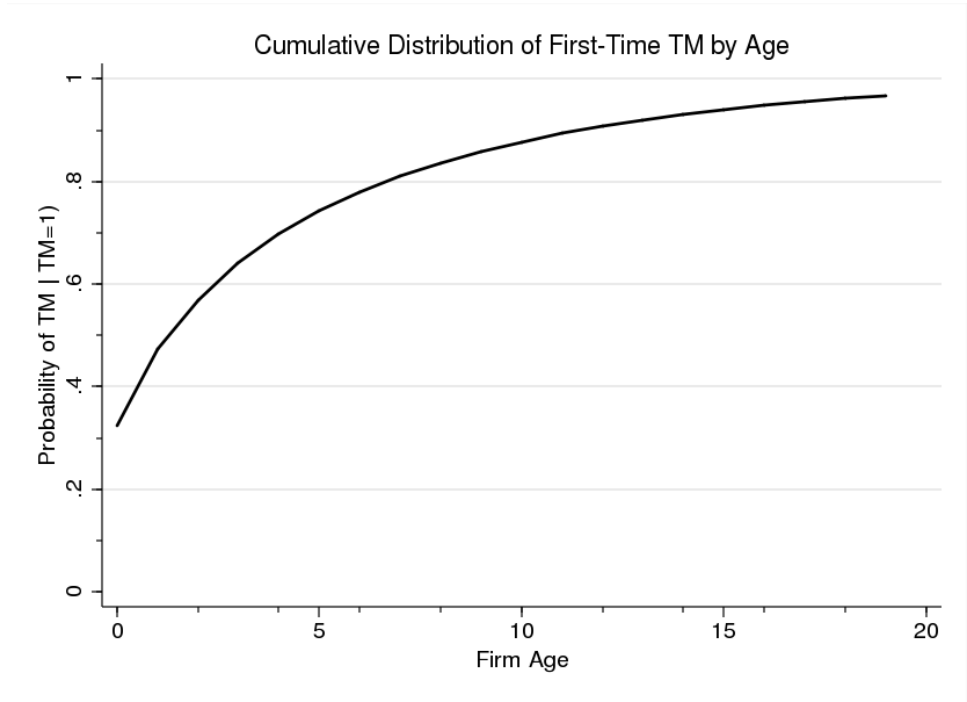


Figure 4: Cumulative distribution function of firm age for first-time trademarkers
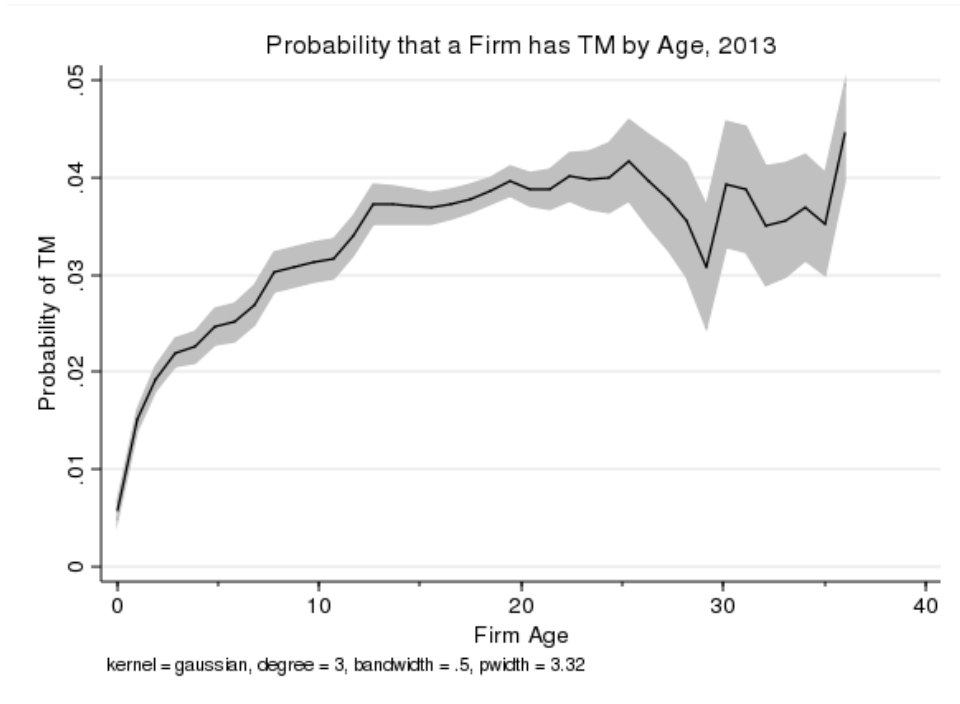
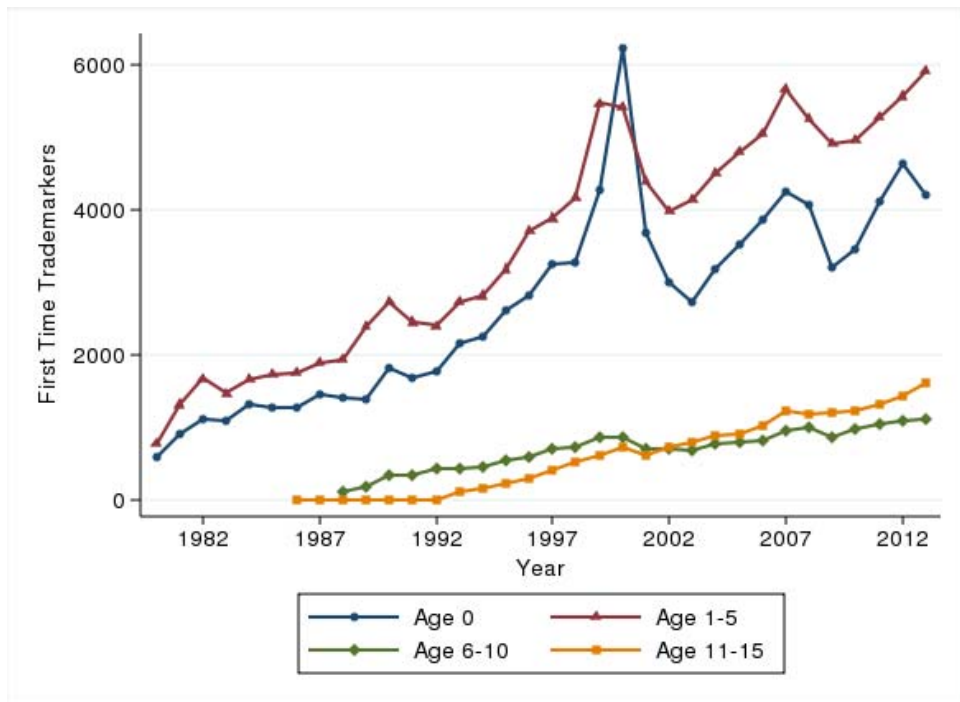Figure 5: Kernel density estimates of firm size (employment) for first-time trademarkers by industry



Figure 6: The age distribution of first-time trademarkers over time
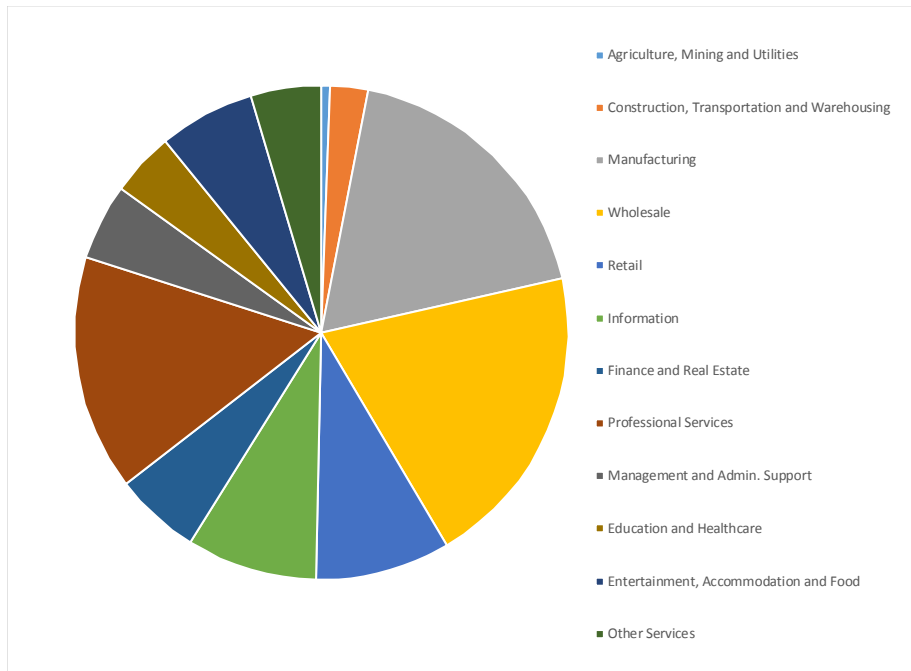
30

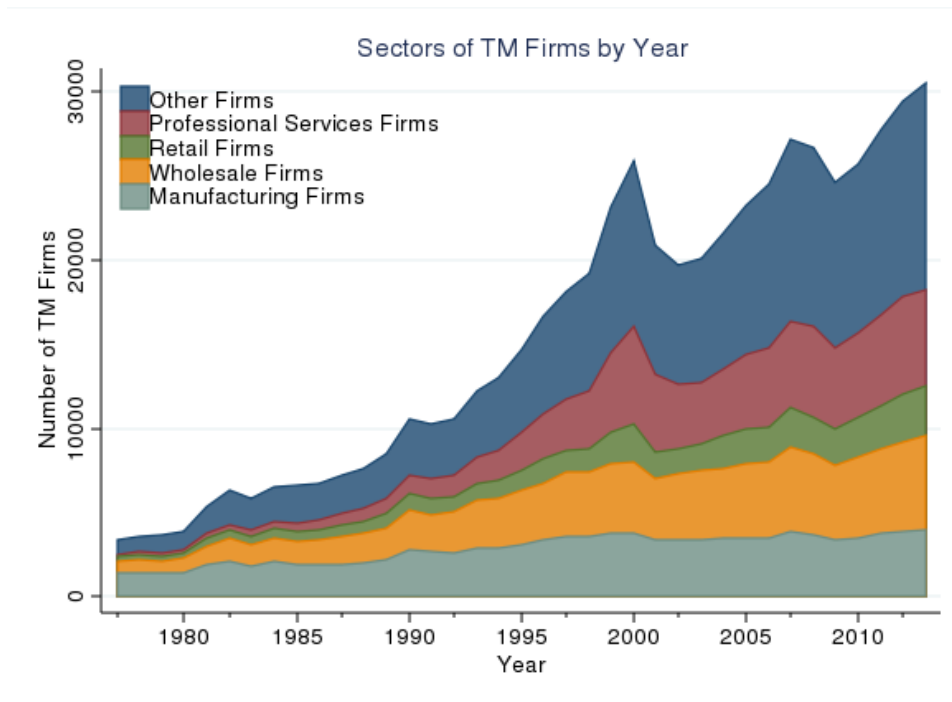Figure 7: The distribution of trademarks across broad industry groups



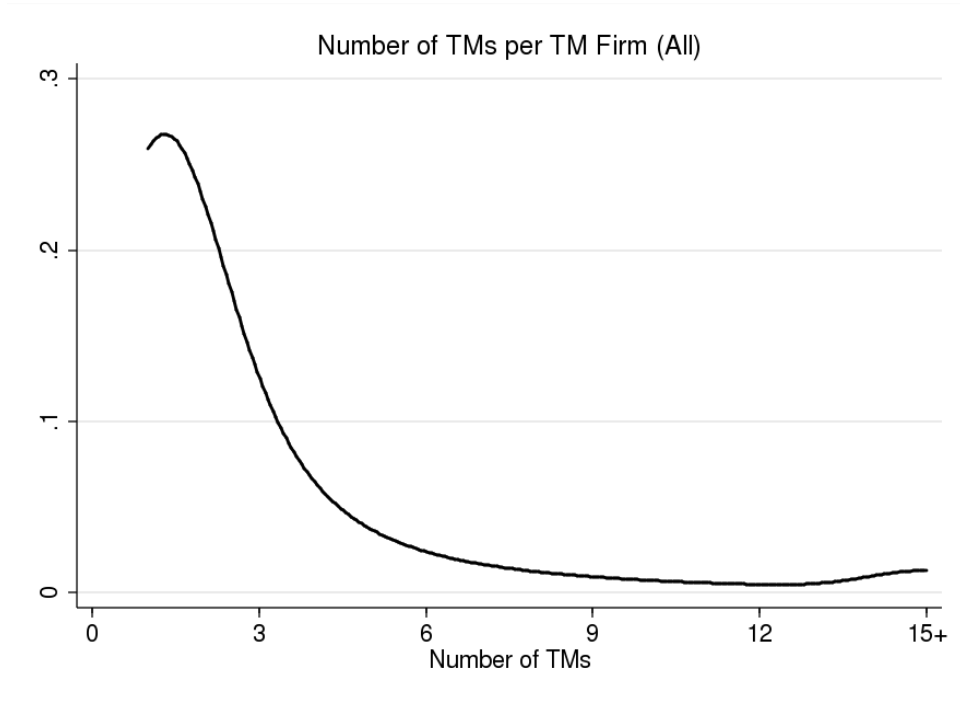Figure 8: The distribution of trademarking firms across industries over time

Figure 9: Kernel density estimate of the number of registered trademarks per trademarking firm
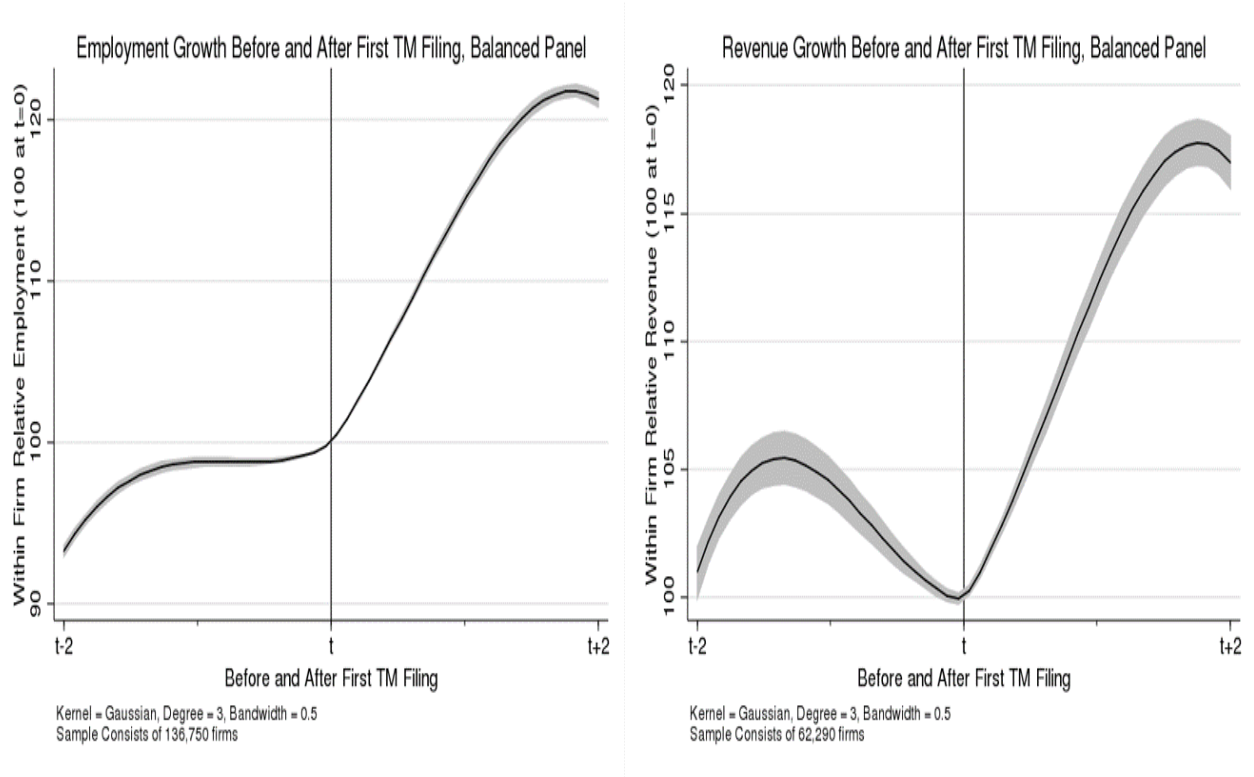


Figure 10: Analysis of the trends in average employment and revenues before and after first trademark filing by a firm
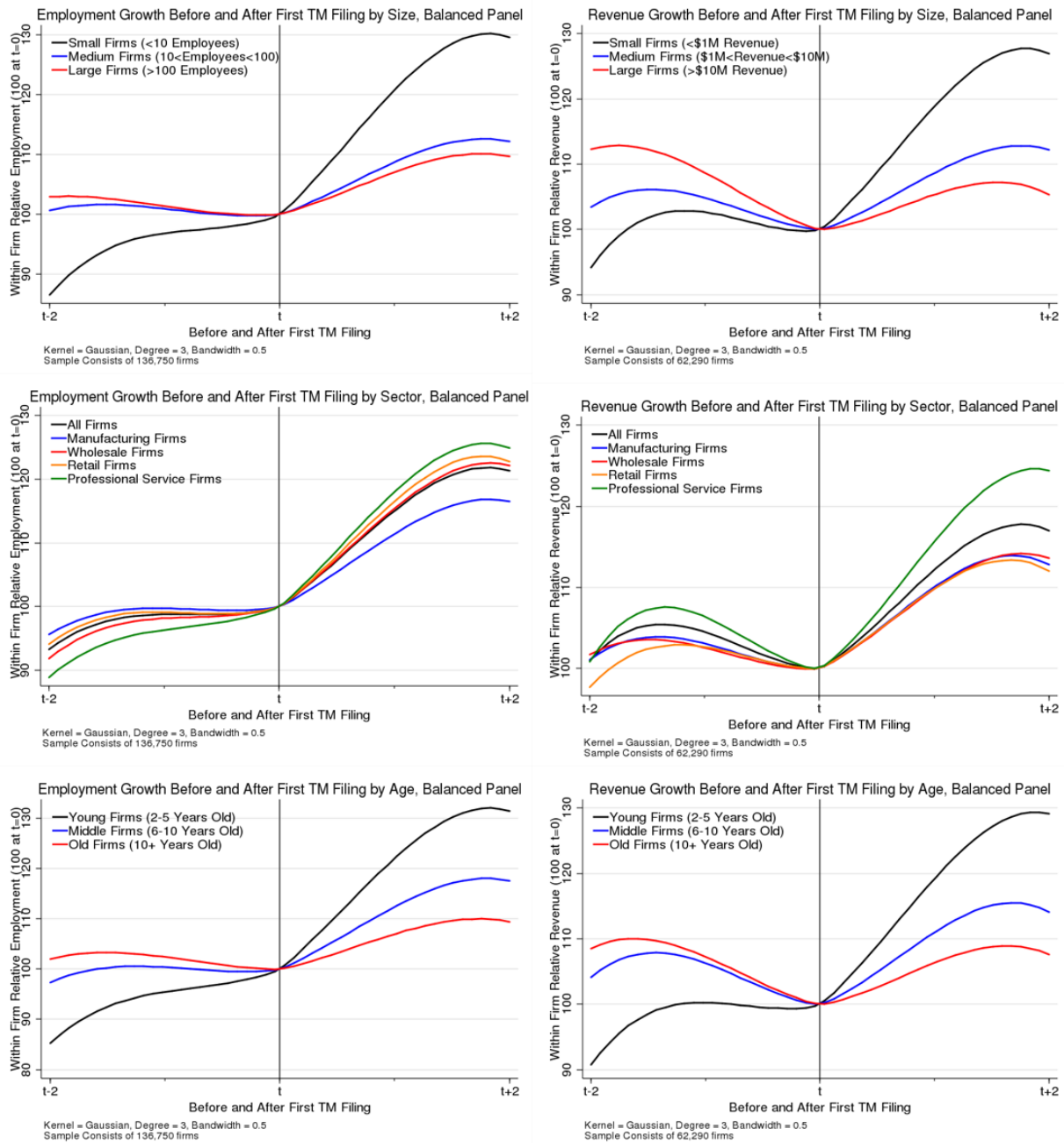
Figure 11: Analysis of the trends in average employment and revenues before and after first trademark filing by a firm - by firm size and age categories, and by sector

Figure 12: The average employment profile of firms that trademark in their first year versus those that never trademark

# Appendix A. Data construction

Because there are no unique identifiers, such as an Employer Identification Number (EIN), that are shared between the TCFD and the BR, the two datasets have to be brought together using name and address matching techniques. The current matching effort is focused on matching domestic businesses observed in the trademark ownership database to the employer firm universe. Future work will incorporate businesses found in the trademark assignments database , which captures the transfer of trademarks between businesses, foreign trademarking firms with domestic establishments, and non-employer trademarking firms.

Several conditions are used to subset the owner file records considered for matching to the BR. First, the country variable is used to identify domestic trademark owners. The owner data file contains owners that are unlikely to be covered in the BR. These include entities representing federal and state governments as well as individuals. The owner entity type code is used to exclude these cases from the match to the BR. Finally, the filing, registration, and publication dates are used to exclude all trademarks for which the maximum year is less than 1976, the earliest year of the BR. The resulting clean TCFD includes all unique domestic name and address combinations observed in the trademarks owner data file. Table A1 shows the count of unique owner records and trademarks in the owner file and after excluding by country and entity type. The owner file initially contains more than 7 million trademarks and 17 million owner records. This reduces to almost 5 million trademarks and 12 million owner records after excluding foreign, federal, state, and individuals and those before 1976.

Table A1. Match frame counts

| Input Frame | Trademarks (000) | Owner Records (000) |
| --- | --- | --- |
| Owner File | 7,214 | 17,381 |
| Excluding pre 1976 | 6,907 | 16,937 |
| Excluding Foreign | 5,939 | 14,406 |
| Excluding Federal, State, Individuals | 5,048 | 12,289 |

Source: USPTO Trademarks Casefile Database, author's calculations.

The first step in the matching process is the extraction of all unique name and address combinations from the owner data file. The owner data file contains an observation for each owner recorded for each trademark application, registration, and publication for years between 1870 and 2015. Each owner record includes the owner's name and address. Moreover, there are multiple name and address types for each record, including "former", "doing business as", and "composed of"names. Each owner record also includes two street address variables, which correspond to the first and second lines of the owner's street address. In many cases the owner's full street address is split across these fields. However, it is not always the case that line 1 should

precede line 2. For example, line 1 might include the suite number while line 2 contains the street address. Therefore, unique names are combined with combinations of line 1 and line 2 of the street address (e.g. line 1, line 2, line 1 concatenated with line 2, and line 2 concatenated with line 1).

The next step in matching the TCFD data to the BR is the cleaning and preparation of both datasets. First, common strings that provide little identifying information for matching are removed from both datasets. These include symbols and punctuation (e.g. " & ", ".", "@"), common words (e.g. "and ", "the"), and abbreviations or designations (e.g. "Co.", "LLC", "LTD "). Additional standardization procedures are used to standardize the name, street, and city fields. Values in the state field are cleaned and standardized and the zip code field is subset to five digits. These cleaning algorithms are applied to the name and addresses from the TCFD and the BR.

The matching of the TCFD to the BR proceeds in several steps. First, an initial subset of potential matches is identified based on a relatively loose name-only match criteria between all unique name and addresses extracted from the TCFD and all establishments in the BR. All matches not meeting this minimum criteria are not considered a match. This initial match is then strengthened using conditions utilizing on both name and address fields. These conditions include nine different match categories, based on whether the match is on name in conjunction with zip code, state or city only, or any combination of them, and name only. Each of these match passes is listed below.

1. Name, street, city, and state

2. Name, city, state, and zip code

3. Name, city, and zip code

4. Name, state, and zip code

5. Name, city, and state

6. Name and zip code

7. Name and city

8. Name and state

9. Name only

The next step of the matching procedure is the cleaning of the matched records. The matches are cleaned and disambiguated using a string comparator . The Jaro-Winkler (JW) string comparator, which takes values ranging from zero and one as a function of how similar two strings are, is calculated for the TCFD and BR name, address, and city pairs. Additionally, a composite JW score is calculated across all three fields. Name-only matches, which tend to be of the lowest quality, are kept only if the JW score between the name fields is greater than 0.90. [23] Across all match passes, only the highest quality pass is kept for each TCFD name and address. Among the remaining matches, the composite JW score is used to select only the highest quality matches. Finally, for matches made using only the name and a single address variable (passes 6, 7, and 8) only matches with a name JW score greater than 0.89 are kept.

The next step of the matching process is aggregating from the establishment -level matches resulting from the BR match described above, to the target firm-level match. All firm identifiers associated with each matched establishment over time along with the first and last year each firm identifier is observed in the LBD are recovered. All unique combinations of TCFD name and address and FIRMID are kept. The first and last firm years, in combination with the minimum and maximum trademarking years associated with each TCFD name and address, are used to clean firm-level matches. Since trademarking activity can plausibly occur well before a firm enters the employer universe, matches outside of five years before and after the firm's first and last year observed in the LBD are excluded.

In order to better understand why TCFD name and addresses do not receive a firm match, unmatched cases are matched to the non-employer business register (ILBD). The ILBD is based on administrative data on income tax returns (Form 1040 with attached Schedule C(s)). In the ILBD some of the observations are individual's names, while others pertain to firm names. To carry out the match, an algorithm is fits applied to determine whether an observation likely represents an individual's name. Steps generally similar to the match in the case of BR are then followed to match the unmatched trademarks to the individuals' names and addresses appearing in the ILBD. At this stage these matches are used primarily to better understand the quality of matches to the employer universe. Future work will investigate matches to non-employer businesses and their growth and transition dynamics.

Table A2 shows the match rates between the TCFD owner file and the BR. In terms of trademarks, the overall match rate to the BR is over 72% with a residual 1.9% of the cases not receiving at least one match to the BR matching to the non-employer universe. The match rates for owner records is similar, with over 74% of records in the owner data file matching to at least one firm in the BR with 1.8% of the remaining records matching to the non-employer universe.

---

[23]This and other cutoff values were reached by visual inspection of the matches and JW scores. The score of 0.90 balanced Type I and Type II errors.

Table A2. TCFD-BR Match rates

| | Trademarks | | Owner Records | |
| | Count (000) | Pct | Count (000) | Pct |
|---|---|---|---|---|
| BR/LBD | 3,736 | 74.0 | 9,163 | 74.6 |
| iLBD | 99 | 2.0 | 223 | 1.8 |
| Unmatched | 1,213 | 24.0 | 2,903 | 23.6 |
| Total | 5,047 | 100.0 | 12,289 | 100.0 |

Source: US Census Business Register, Integrated Longitudinal Business Database, USPTO Casefile Database, author's calculations.

Notes: Counts rounded to avoid the disclosure of sensitive information.

These rates include cases where a single owner record or trademark matches to more than one FIRMID. The analyses in this paper only use cases where a unique FIRMID can be identified for a given trademark. Table A3 shows the share of trademarks and owner records that receive a single matched FIRMID. Almost 54% of trademarks receive a unique match and over 84 % received three of fewer matches. These patterns are similar for owner records. Over 56% of owner records are matched to a single FIRMID and almost 87% receive three or fewer matches.

Table A3. TCFD-BR Match ambiguity

| Firm Matches | Trademarks Count (000) | Pct | Owner Records Count (000) | Pct |
|---|---|---|---|---|
| 1 | 2,020 | 54.1 | 5,158 | 56.3 |
| 2 | 760 | 20.3 | 1,914 | 20.9 |
| 3 | 379 | 10.2 | 896 | 9.8 |
| 4 | 199 | 5.3 | 448 | 4.9 |
| 5 | 123 | 3.3 | 263 | 2.9 |
| 6 | 87 | 2.3 | 189 | 2.1 |
| 7 | 54 | 1.4 | 105 | 1.2 |
| 8 | 37 | 1.0 | 67 | 0.7 |
| 9 | 25 | 0.7 | 47 | 0.5 |
| 10+ | 51 | 1.4 | 76 | 0.8 |
| Total | 3,736 | 100.0 | 9,163 | 100.0 |

Source: US Census Business Register, USPTO Casefile
Database, author's calculations.

Notes: Counts rounded to avoid the disclosure of
sensitive information.

There are a number of trademark and owner record characteristics that can be used to better understand the TCFD-BR match. Table A4 shows the owner record match rate by owner record type, which captures the point in the trademark life cycle when the owner was recorded. The match rates are stable across different owner record types. Over 72% of the 5 million filer records receive at least one match, which increases to 83% for transfer records.

Table A4. Match rates by owner record type

| Owner Type | Total Count (000) | Match Rate |
|---|---|---|
| Filer | 5,009 | 72.6 |
| Publisher | 3,732 | 75.4 |
| Registrant | 3,370 | 76.2 |
| Transfer | 177 | 82.9 |

Source: US Census Business Register, USPTO Casefile
Database, author's calculations.

Notes: Counts rounded to avoid the disclosure of
sensitive information.

Similarly, Table A5 shows the match rates by trademark entity types. Perhaps not surprisingly, corporations have the highest match rate ( over 82%). Sole proprietors have the lowest match rate with only 16% of records receiving at least one match.

Table A5. Match rates by owner entity type

| Status type | Total Count (000) | Match Rate |
|---|---|---|
| Corporation | 9,170 | 82.1 |
| Partnership | 168 | 42.2 |
| Sole Proprietorship | 68 | 16.4 |
| Other | 2,837 | 54.0 |

Source: US Census Business Register, USPTO Casefile Database, author's calculations.

Notes: Counts rounded to avoid the disclosure of sensitive information.