NBER WORKING PAPER SERIES

MEASURING THE IMPACT OF HOUSEHOLD INNOVATION USING ADMINISTRATIVE DATA

Javier Miranda
Nikolas Zolas

Working Paper 25259
http://www.nber.org/papers/w25259

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2018

I Javier Miranda attest that I have not received any funding for this work outside of my regular salary at the U.S. Census Bureau. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Measuring the Impact of Household Innovation using Administrative Data
Javier Miranda and Nikolas Zolas
NBER Working Paper No. 25259
November 2018
JEL No. O3,O31,O35

ABSTRACT

We link USPTO patent data to U.S. Census Bureau administrative records on individuals and firms. The combined dataset provides us with a directory of patenting household inventors as well as a time-series directory of self-employed businesses tied to household innovations. We describe the characteristics of household inventors by race, age, gender and U.S. origin, as well as the types of patented innovations pursued by these inventors. Business data allows us to highlight how patents shape the early life-cycle dynamics of nonemployer businesses. We find household innovators are disproportionately U.S. born, white and their age distribution has thicker tails relative to business innovators. Data shows there is a deficit of female and black inventors. Household inventors tend to work in consumer product areas compared to traditional business patents. While patented household innovations do not have the same impact of business innovations their uniqueness and impact remains surprisingly high. Back of the envelope calculations suggest patented household innovations granted between 2000 and 2011 might generate $5.0B in revenue (2000 dollars).

Javier Miranda
U.S. Bureau of the Census
Center for Administrative Records Research
4600 Silver Hill Road
Washington, DC 20233
javier.miranda@census.gov

Nikolas Zolas
Center for Economic Studies
U.S. Census Bureau
4600 Silver Hill Road
Washington, DC 20233
nikolas.j.zolas@census.gov

1 Introduction

The study of innovation has traditionally centered on the institutions where it is believed to be conducted, which has primarily consisted of the firm. The underlying assumption is that innovation is the output from an R&D production function that has the inventor at its core and where the inputs (materials and human capital) are fully accounted for. Some of the inputs may take the form of knowledge originating outside the firm like universities, government labs and other firms. In this regard government and universities labs have long been recognized as sources of knowledge and invention. Other firms may contribute to the R&D process through research joint ventures or may license their technologies. Increasingly however, researchers are highlighting the importance private households play as sources of invention and innovation in this process (e.g. von Hippel, de Jong, and Flowers (2012) and Arora, Cohen, and Walsh (2016)). In this paper we aim to contribute to this strand of the literature by using U.S. Census Bureau administrative data combined with USPTO patents data to document household innovations. The use of administrative data gets around some of the problems with current studies in this area, specifically small sample sizes in household surveys, low power estimates, and low response rates which may raise questions about non-response bias (Deming, 1990).

Use of administrative data provides a rich picture of the types of innovations undertaken by households and their characteristics but it has its own limitations. We focus on the set of household innovations we can identify in administrative data; that is, those that are granted a patent by the USPTO. Admittedly this excludes perhaps what might be the lion share of household innovation; that which is not patented. By contrast we focus on what might be the most valuable innovations (Arora, Cohen, and Walsh, 2016) and we do so in a systematic manner. We match these patents to Census Bureau administrative files to understand the demographic characteristics of household inventors as well as the characteristics of the unincorporated businesses they start to get a sense for their impact and value.[1] Use of administrative records comes with other important limitations. Specifically, there is no way for us to determine whether these patents where developed during leisure time or as a remunerated activity. Here we make the strong assumption that if they have not been assigned to a firm there was no direct remuneration for the development of the innovation.

When documenting the characteristics of household innovations we describe the technology classes they fall under, their impact and novelty as captured by the analysis of backward and forward looking citations, and the breadth of their application as captured by a generality index. In addition we document the characteristics of inventors, their age, gender, race and origin. When looking at business formation we examine the dynamics of unincorporated businesses that are tied to inventors and their performance relative to similar businesses without inventors specifically their revenue and growth performance.

We find household inventors are disproportionately U.S. born relative to salaried inventors. They are relatively white. Household inventors are disproportionately under 25 and over 55. Across the board we find a deficit in female and black inventors relative to the overall working age population.

Household inventors work on technology classes disproportionately tied to consumer products such as Design, Mechanical and Other. Patents associated with household innovation are about 1/2 as likely to be considered "radical".[2] In terms of value, household innovations accumulate approximately 27-33% fewer citations on average. While their citation impact is smaller it remains remarkably high. Finally, we find few household inventors attempt to create a business around their invention. When they do these businesses have higher revenues on average and are more than twice as likely to transition to hire their first employee than nonemployers who do not patent. Back of the envelope calculations suggest

[1] Patents by independent inventors have been found to display the largest rates of transfer (Serrano, 2010) so in future drafts we will explore the characteristics of patents that transition to existing firms.

[2] We follow the definition in (Dahlin and Behrens, 2005), a radical innovation is one that is considered novel, unique and impactful.

patented household innovations granted in a given year might generate revenue $5.0B in revenue in 2000 dollars.

The remainder of the paper is structured as follows. Section 2 provides background. We follow with a description of the data in section 3. We describe basic features of patented household inventions in section 4. Our analysis of business formation and outcomes follows in section 5. We conclude in section 6.

2 Background

Innovation is traditionally thought of as a process that takes place inside of a firm. In this context, outside sources of knowledge and invention including universities, government labs and other firms have long been recognized as important inputs to the firm's R&D function. Increasingly however innovation researchers are focusing on households as important sources of knowledge and innovation. The study of household innovation however has been hampered by data availability.

The first set of household innovation studies looked at user innovations in specific product markets. Early examples include von Hippel (1976) and Shah (2000) looking at user innovation in scientific instruments and new sporting goods respectively. Their methodology involves a retrospective study of a selected sample of commercially successful innovations as identified by either experts in the field or by direct analysis of new product features. This was followed by interviews of relevant product and industry experts. Both these authors find a large percentage of the innovations were in fact invented, prototyped and tested by users of the equipment rather than the equipment manufacturer. In the case of scientific instruments, von Hippel (1976) finds existing instrument manufacturers would incorporate user innovations into their products with a focus on improved engineering. In the case of sporting goods Shah (2000) finds users built innovative equipment for their own use. The inventors tended to be young and they often built businesses in order to appropriate the benefits from their innovations.

Follow-up studies have tried to more broadly describe the characteristics of the innovators and rate of user innovation. Lüthje (2004) conducts a survey of users of outdoor sporting equipment identified from the direct mail order listing of two sporting goods manufacturing firms. While response rates are relatively low at 26%, the author finds a large share of respondents, 37%, claimed at least one idea. Of these 30% claimed their idea provided a solution to a problem that was not offered by the manufacturer. Reportedly, only 4 in 10 took their ideas beyond concept by developing prototypes. Franke and Shah (2003) look at innovation within four distinct communities of extreme sports enthusiasts. Communities of consumer users were identified through websites or competition rosters. With a survey response rate of 38% the authors find 32% of community members claimed an innovation and of these 14.5% considered the innovation to be a completely new product. In their sample, 23% of innovators believed their innovations had been or would be commercialized by a third party. These innovators did not appear to benefit financially from their innovations. Whether results from these and other surveys of lead users and enthusiasts are representative of broader user communities remains an open question.[3]

von Hippel, de Jong, and Flowers (2012) take a broader approach to this question by conducting a household survey to look at inventions by a representative sample of consumers in the U.K. These are innovations tied to households and their unincorporated businesses. Specifically they look at the development and modification of consumer products by product users. The types of household innovations they focus on exclude on-the-job innovations which are already accounted for in official statistics. Instead they focus on innovations that were developed during uncompensated leisure time. With a survey response rate of 15% they find 6.2% of U.K. consumers engaged in consumer product innovation in the previous 3

[3] A good survey of consumer user studies can be found in de Jong (2016).

years. When comparing against the amount of R&D investment by U.K. firms they estimate the volume of household based expenditure exceeded that of firms by a factor of 2.3 times.[4] They conclude private households are a major source of invention.

In a survey centered on consumer product innovations von Hippel, de Jong, and Flowers (2012) find 98% of innovations are product modifications rather than new product creations. Most of the innovations, 80%, are in a few product classes that are related with how people spend their time: craft and tools, sports and hobbies, gardening as well as child, dwelling or pet related. Only 17% of the innovations are believed to be adopted by others to some degree and only 2% of the innovations are protected by intellectual property rights. There are relatively few software innovations. von Hippel, de Jong, and Flowers (2012) are the only study collecting demographic information from a representative consumer sample rather than a community of interest. They find inventors tend to be male, educated, and either a student or age over 55. Issues with this and other representative consumer surveys that have followed include high non-response rates, small sample sizes, and confusion regarding the definition of innovation by consumers. With these limitations a general conclusion is the apparent low adoption rates of innovations by enterprises.

Following a different approach Arora, Cohen, and Walsh (2016) conduct a survey of manufacturing firms to examine the extent to which firms in the U.S. use external sources of invention for their innovations. Arora, Cohen, and Walsh (2016) focus on the whole manufacturing sector regardless of industry or whether firms own patents or conduct R&D. Their sample is drawn from the Dun & Bradstreet business frame but adjusted with U.S. Census Bureau based weights to match the population of manufacturing firms by industry, size and age. For the analysis they focus on product innovations (and exclude process innovations) at firms with more than 10 employees. With response rates of 30.3% they find that of the 16% of firms that innovated (introduced a product that is new to the market),

[4]von Hippel, de Jong, and Flowers (2012) find the average customer invention requires an expenditure of £101 and 4.8 days.

49% report their most important new product originated from outside. They find customers are the most pervasive source of invention although not the source of the most valuable ones. The more valuable inventions are sourced from technology specialists which include independent inventors. These inventors patent their own inventions at relatively high rates, 56%, higher than university, supplier and customer sourced inventions at 36%, 34% and 16% respectively. They find independent inventors are also a more common source of inventions for small firms.

3 Data

We focus our analysis on patented household innovations. Our primary source of patent data is the U.S. Patent and Trademark Office PTMT Custom Patent Data Extract. These data are produced annually from the bibliographic text (i.e., front page) of the patent documents. It covers all granted patents by the USPTO and detailed information including the patent number, type of patent, filing date, issue date, inventor information, assignee name at time of issue, and classification information for each.

We use the patent class information to impose some initial restrictions on the patents we analyze. Depending on the patent documents, patents can be assigned to firms, individuals or governments. These can each be either domestic or foreign. In addition the patents can be unassigned. This happens when the inventors have not granted the rights to the invention to a corporation, university or government agency, or to other individuals. In these cases the patents are assumed to remain with the inventor, but in some cases can later be reassigned to firms. We exclude from the set of patents we analyze those that belong to governments and all foreign patents. We assume these are not tied to independent U.S. based inventors. Table 1 looks at the number of patents by assignee type in our sample. We center our analysis on patents granted between 2000 and 2011. Our sample includes a total of 1.29 million patents granted between 2000 and 2011. The bulk of these, 80%, are assigned to businesses. Most of the remaining patents, 19.2%, are unassigned. There are very few patents, 0.8%, assigned to individuals. While unassigned patents are assumed to belong to the inventor it will be the case that some of these belong to firms but were not assigned at time of grant. We explore the extent of this problem by reviewing patents with large team sizes of inventors to get a sense for the amount of noise in the data. Our assumption is that the average firm patent will be developed by larger teams of inventors. The results can be seen in Figure 1. The team size distributions for unassigned and individual assigned patents are fairly similar and well to the left of firm assigned patents. Unassigned patents have the larger share of single inventor patents (nearly 80% of unassigned patents have a single inventor). Looking at the right tale of the distribution, we find that fewer than 1% of unassigned and individual assigned patents have inventor team sizes of 5 or more, compared to the nearly 7% of firm assigned patents.

Firm assigned patents present a challenge to us. The patent data does not include firm identifiers or flags that might help us distinguish patents assigned to employers from those assigned to nonemployer businesses. It is not unreasonable to think however that independent inventors might assign their patents to their own unincorporated nonemployer business. However, we do not want to exclude these inventors from our analysis since their patents might be particularly valuable. We rely on the U.S. Census Bureau longitudinal patent-business database (BDS-IF) to identify and exclude from our analysis patents assigned to employer businesses while keeping those assigned to nonemployer businesses.[5] We identify patents assigned to nonemployer firms by matching all patents to the U.S. Census Bureau's Business Register of nonemployer firms.[6] A large percentage of patents, nearly 80%, match to the employer universe files. The employer matches tend to be based on the assignee

[5] The BDS-IF identifies patents assigned to employer businesses while keeping those assigned to nonemployer businesses. See Graham, Grim, Islam, Marco, and Miranda (2018) for details of the matching methodology. Briefly, it uses both the assignee and inventor information to form a match. The use of two independent pieces of information to identify the assignee firm provides a high level of reliability in the match.

[6] All businesses that file an income tax form to the IRS authorities and have no associated payroll tax form are included in the nonemployer Business Register. See Appendix A for details of the matching methodology.

name and address, while the nonemployer matches mostly occur through the inventor. We remove the known employer matches from Graham, Grim, Islam, Marco, and Miranda (2018) from our universe of matches, leaving us with approximately 200,000 raw nonemployer firm matches. Our set of initial matches require further refining. A high quality firm-inventor match does not guarantee the inventor is matched to its firm. Think of an inventor named David Smith in D.C. and a company named David Smith located also in D.C. First, there are possibly many unincorporated entities named David Smith so the match might not be unique. Even if the match is unique we do not know whether the owner is the inventor (i.e. there are many David Smiths). We retain only cases where the social security number of the inventor and the social security number in the nonemployer firm record line up.[7] This filtering process leaves us with a total set of approximately 125,000 patents. We remove an additional 55,000 patents by only keeping the unduplicated matches.[8] Finally, we drop patents that are associated with nonemployers that have an unusually large number of patents assigned to them.[9] This leaves us with a total of 68,000 patents associated to nonemployer businesses that we are confident belong to the inventors behind the patents.

Table 2 shows the percentage of patents matched to employer businesses (E) and nonemployer business (NE) by assignee type and year. Patents that remain unmatched (U) are not associated with business activity as captured by the Business Register. To be clear,the set of unmatched patents might include some that were matched to a Census dataset but where the match either could not be uniquely resolved to an inventor or assignee or was not linked through the PVS process.[10] Table 2 highlights a clear separation in the match

[7]This comparison is done indirectly. The Census Bureau strips personally identifiable information from all of its internal files to protect the confidentiality of records. Specifically the Census Bureau replaces and individual's name, address (and SSN if present) with a Protected Identification Key (PIK) using the PVS system (Wagner and Layne, 2014). Each name-address pairing has a unique PIK in the system. The Census Bureau assigned a PIK to the patent data using the name and location information.

[8]The PVS system does not guarantee an inventor in the USPTO database will receive a unique person identifier. In cases where the identifying information is not unique enough multiple PIKs are assigned.

[9]These might be holding entities with no associated employers.

[10]See Appendix A for details of the match and Table A8 for a breakdown of unmatched patents

rates by assignee type, with the vast majority of firm-assigned patents linked to employer firms. By contrast individual-assigned patents have much lower match rates. Only about 50% of patents are associated with some form of business activity with most of it tied to nonemployer firms. Only 30.4% of unassigned patents are tied to some form of business activity. We have no links to employer businesses since there is no available firm assignee information in the patent document for these patent types.

4 Characteristics of Patented Household Innovations

In this section, we describe the characteristics of patents and inventors associated with what we call patented household innovations which include patents that are either unassigned or are assigned to individuals. We contrast those with patents assigned to firms. We start by describing differences in the demographic composition of the inventors associated with the patents, before delving into the characteristics of the actual patents.

4.1 Inventor Demographics

To highlight potential differences in demographic characteristics of inventors associated with household innovations we link demographic information from administrative U.S. Census Bureau files to the inventors in the patent records. They provide basic demographic information including gender, race, country of origin and birth date for all people in the U.S. with a social security number.

Information from the demographic files is linked by use of a Protected Identification Key (PIK) available on both sets. We are not able to uniquely identify all inventors in the patent documents in our files due to limitations of the data.[11] There are 1.48M inventors associated with the 1.29M patents that form our analysis. We are able to obtain demographics for

[11] The identification would be greatly facilitated if the USPTO were able to collect either a birth date or a SSN/TIN.

inventors on 856,000 of the 1.29M patents.[12] Overall we find inventors tied to firm assignees are more likely to be uniquely identified than individual assignees or unassigned patents. We also find that patents unmatched to demographic data are mostly concentrated in the sectors of "Design" and "Plants". Details of the matching procedures are results can be found in Appendix B.

Table 3 shows demographic information for the set of inventors we were able to identify by assignee type and type of economic activity. There are some notable differences in the demographic composition of the patent types but also some similarities. The first thing to notice is that the vast majority of patents are filed by males. This is true across all assignee types and is consistent over time.[13] Innovation activity whether household or firm based is a male dominated activity. This is consistent with Bell, Chetty, Jaravel, Petkova, and Van Reenen (2016), who find a similar deficit in female innovators.

Firm-based patents disproportionately favor foreign-born inventors relative to individual-assigned patents and unassigned patents, with approximately 1/3 of inventors affiliated with firm-assigned patents being foreign born, compared to 20% for other assignee types. Given this it is perhaps not too surprising that firm-assigned patents are less likely to be associated with black or white inventors and nearly twice as likely to be associated with "other" races relative to individual-assigned and unassigned patents. The share of foreign born inventors outweighs their relative share in the labor force at 16.7% of the total in 2015.[14] We find there is a deficit of black inventors across the board again consistent with Bell et al. (2016).[15]

Finally, individual-assigned and unassigned patents disproportionately favor both older (over 55) and younger inventors (less than 25). Nearly 1/3 of the household inventors are 55 years and older, compared to the 20% found in firm-assigned patents. This is consistent

[12] We are able to identify demographics from 884,000 patents, but 28,000 of the patents are later classified as reassigned, which are dropped from our analysis

[13] Time series results not shown.

[14] Shares of foreign born in the labor force are reported in Bureau of Labor Statistics (2016).

[15] Blacks and whites made a 12% and 79% respectively of the labor force population in 2015 (Bureau of Labor Statistics, 2016).

with von Hippel, de Jong, and Flowers (2012) who find household innovations are disproportionately tied to students and men over 55.

To summarize our findings, household innovators (associated with individual assigned and unassigned patents) are more likely to be US born, white, less than 25 and over 55 than firm based innovators. In the case of the latter, the proportion of household innovators above the age of 55 is more than 12 percentage points higher (31.6 versus 18.8). Across the board we find a deficit of female and black inventors relative to the population of employed workers and over representation of foreign born inventors.

4.2 Technology Class

We next focus on the types of technology classes associated with household innovations. Previous research has focused on consumer product innovations and found innovations tended to be focused in a few product classes. Here we focus on the broader set of patented innovations. We look at the technology composition by assignee-type. We also look at those that lead to direct business activity and those that do not. For our classification, we use the primary USPC code assigned to each patent and group them into eight broad classes consisting of: Chemicals, Computers and Communication (C&C), Design, Drugs and Medicine (D&M), Electrical and Electronics (E&E), Mechanical, Plant Patents and Other. The grouping by USPC class is based on Hall, Jaffe, and Trajtenberg (2001) and expanded to include new patent classes as detailed in Dreisigmeyer, Graham, Grim, Islam, Marco, and Miranda (2014). Table 4 shows the breakdown by assignee type. We find firm assigned patents are disproportionately in Chemical, C&C and E&E relative to individual assignee and unassigned patents. By contrast they are underrepresented in Design, Mechanical and Other. Table A12 in Appendix C provides a listing of technology subcategories associated with each broad class. Amongst the technologies included in Mechanical and Others are Motors, Engines & Parts, Transportation and Miscellaneous such as hardware and tools. Others include Amusement Devices, Apparel & Textile, and Furniture & House Fixtures and miscellaneous such as Robots and Aquatic Devices. All fairly typical consumer products. Design patents provide protection to ornamental designs embodied in or applied to an article of manufacture. Analysis of the top 50 companies having been granted design patents shows that these are dominated by technology, automotive, and consumer product companies.[16]

Table 5 breaks down the previous table by business activity. The patterns here replicate the findings discussed regardless of business type. A few things stand out. First, the majority of Design patents are not associated with business activity and remain unmatched. This is true for both individual assigned and unassigned patents and suggests fundamental differences perhaps in the value of design patents vis-a-vis utility patents and maybe the requirements for grant. Second, patents with a firm assignee in the Drugs & Medical class are harder to match to business databases perhaps due to the complex structure of firms developing them.

We combine our technology classes with the individual demographics to identify compositional differences between employer patents and household innovations. Table 6 takes the difference in the proportion of patents by technology class and demographic characteristic between nonemployer patents and employer patents. The table highlights several key differences, most of which are significant. Design patents clearly differentiate themselves in terms of demographics. The previous sections have alluded to the fact that nonemployer patents are disproportionately male, U.S. born, white and older than employer patents. However, this does not seem to be the case for Design patents, where the opposite holds. It appears design patents in employer businesses are disproportionally associated with white, male, U.S. born inventors where they might hold a relative advantage signaling to the very different nature of these types of patents.

[16]For details see report from Intellectual Property Owners Association (2015).

4.3 Team Size

Evidence from surveys and product studies suggest the complexity and knowledge embodied in household innovations might not run very deep. A typical story might be that of a consumer that modifies the face of a clock to teach their kids how to tell time.[17] Consistent with this survey data also shows that the average expenditure in developing a household innovation is not very high. In this section we explore whether this is also true of patented household innovations. We follow Jones (2009) and use team size as a measure of the complexity and depth of knowledge associated with a particular innovation. The burden of knowledge hypothesis would indicate household innovations require smaller team sizes.

Figure 1 plots the distribution of team sizes by assignee types and shows that firm-assigned patents tend to be significantly larger on average. The size distribution for individual assignee and unassigned patents is fairly similar and rests well to the left of firm assigned patents. A large share of individual-assigned and unassigned patents are developed by a single inventor relative to patents assigned to firms. There are single inventors on 60.7% of individual-assigned patents and 83.5% of unassigned patents versus 30.8% on firm-assigned patents. Table 7 tabulates the mean team size by assignee type, technology class and type of business and finds similar results across them. Team sizes for patents matched to nonemployer business tend to be significantly smaller on average than patents matched to employer firms, having on average nearly one less team member. Patents with no associated business activity have the smallest team size on average. Consistent with Jones (2009) and Kim and Marschke (2015) Drugs and Medicines and Chemicals tend to be composed of the largest inventor teams, while Design patents consist of the smallest teams.

[17] This story is taken from von Hippel, de Jong, and Flowers (2012)

4.4 Impact

Household survey data indicates the impact and quality of household innovations might not be very high. Survey respondents often indicate they do not expect their inventions to be adopted. In this section we explore whether this extends to patented household innovations. In this section we follow the literature and use citation counts as a noisy measure of the quality of a patent and their technological impact. We then use a new measure of impact that takes account of the structure of forward and backward looking citations to identify radical patents. Finally we examine whether these innovations are general purpose or instead narrow in application. We ignore truncation issues in the analysis assuming similar impacts across types of patents.

4.4.1 Citations

For our citation measures we use the latest citation count (as of December 2015) collected from PatentsView and link them to our dataset. Figure 2 shows the distribution of citation counts by assignee type. Table 8 reports the means by assignee type, business type and broad technology class. On average, individual-assigned patents have a lower mean citation count than firm-assigned patents. The mean citation for firm-assigned patents is 16.4, while the mean citation count for individual-assigned patents is 11.3 and 10.2 for unassigned patents.[18] The difference in average citation counts is driven in part by an across the board lower citation count across technology classes. However, some of the largest differences in mean citation counts can be found in the Design, Mechanical and Others categories precisely the areas where household innovations are concentrated so composition effects contributes to the overall difference. More interestingly perhaps is the finding that household innovations are quite heavily cited on average and in some areas such as Computers & Communications

[18] Approximately 160,000 patents out of the 1.29M have zero citations. The proportion of patents with zero citations by matched data and assignee type is approximately equivalent to the proportion of total patents by matched data and assignee type.

and Electrical & Electronic the difference is not very large. Looking at the citations across type of business activity we find patents have a mean citation count of 16.4, 13.4, and 11.4 respectively for patents associated with employer businesses, nonemployer businesses and no business activity. Again, these differences appear to be driven by composition effects as well as generally lower citation counts within particular technology classes.

To examine differences in citation counts after controlling for technology composition we run a Poisson regression on citations looking at the impact of business type after controlling for patent class (main USPC code) and grant year. The results are found in Table 9. Column (1) looks at citation impact by business type and Column (2) by assignee type. Focusing on Column (1) we see the difference in the logs of expected citations is 0.288 units higher for patents matched to employer firms and 0.06 units higher for patents matched to nonemployer firms relative to unmatched patents, holding everything else constant. This is equivalent to a citation count that is 33.4% higher for employer-matched patents and 6.2% higher for nonemployer-matched patents, for a difference of 27% in citations between employer and nonemployer patents. Looking at the differences in citations by assignee type, Column (2) we find a similar difference between firm assigned patents and individual assigned patents. The coefficient values give a difference in the logs of expected citations to be 0.096 units higher for firm assigned patents and -0.247 units lower for individual assigned patents relative to reassigned patents. This is equivalent to a citation count that is 10% higher for firm assigned patents and 22% lower for individual assigned patents, for a difference of 32%.

4.4.2 Radical Patents

Households innovators will be relatively resource constrained compared to firms. These innovators might choose to focus on technologies that require smaller investments and prior knowledge –they are not complex. Consistent with this idea section 4.2 documented the disproportionate weight design patents have amongst household innovators. In this section we explore whether this might lead them also to work on innovations that represent breaks with past knowledge within specific technology fields. In this section we assess the proportion of breakthrough patents amongst patented household innovations as defined by whether they represent a "radical" break from existing knowledge in that field. Since it is the focal point of a new technological trajectory the patent itself must be cited.

Our measure builds on the concepts of Dahlin and Behrens (2005) but is extended to the universe of patents in the USPTO patent database (Dreisigmeyer et al., 2014). Dahlin and Behrens (2005) define the term radical invention as one that meets three properties: 1) it is novel - it has distinctive features that are missing in previously observed inventions; 2) it is unique - it is the focal point of a new technological trajectory; 3) it must be adopted - it should influence future inventions. The authors operationalize this idea by examining both forward and backward citation patterns for any given patent. Forward citations are citations to a patent made by other later patents. It is a measure of the patents impact on future inventions and its value in the market. Backward citations are defined by the prior art cited by the patent itself. Backward citations contain information about the radicalness of the innovation. The more radical a technology the more likely it is to cite prior art outside its own patent class since this will necessarily involve combining different elements rather than inventions from its own field.

Table 10 reports the number of patents (per thousand) that qualify as being radical by assignee type, business type and technology class. In general, patents matched to employer firms are more than twice as likely to be considered radical versus patents matched to nonemployer firms and unmatched patents. This does not appear to be driven by compositional differences in the patent types, as employer-match patents and firm-assigned patents consistently have higher rates of radical patents across all technology classes. Design patents appear to have high rates of radical innovations. Many of these appear to be self referencing and not have much of an impact outside the patenting firm suggesting these might be disproportionately defensive patents. While there are relatively fewer radical patents amongst household innovators there is still a non trivial number of them. We examine some of the radical patents identified. The bulk of them are found in Computers & Communications, Design, and Drugs & Medical. They include a system for providing traffic information to a plurality of mobile users connected to a network, a system for dynamically pushing information to a user utilizing global positioning system, a method and apparatus for securing a suture and a flash memory drive with quick connector. All these technologies had broad impacts in their fields.

4.4.3 Generality Index

Finally we describe the breath of impact patented household innovations have outside of their own field. Some technologies are more specific with a limited application across industries while others have a wider field of application. We use the patent classification codes to generate a measure of generality, G_i, that is close to that used by Hall and Trajtenberg (2004) as follows:

$$G_i = \sqrt{\sum_j^n s_{ij}^2}$$

where s_{ij} denotes the percentage of citations received by patent i that belong to patent class j, out of n patent classes. This is simply the square root of the Herfindahl concentration index and therefore if a patent is cited by subsequent patents that belong to a wide range of fields the measure will be low and close to zero. By contrast if the citations are concentrated in a few fields the measure will be close to 1. Further, if a patent has a single citation in the same technological field this measure will be equal to 1 and it won't be defined when it receives no citations.[19]

We compute a Generality Index for patents in our sample that were granted up through 2008 to limit the impact of right censoring. Table 11 looks at the mean generality index

[19] This modified measure of generality retains important properties of metric spaces (or distance functions) that allow us to measure the distance, instead of just a similarity, between two patents.

by assignee type, type of business activity and technology class. In general firm assigned patents find application across a broader set of technological fields. This is particularly true for Chemical, Drugs & Medical and Mechanical. Independent inventors appear to focus on technologies that have narrower impacts. Across the board and as expected patents in Computers & Communications and Chemical have broader applicability receiving the highest number of citations outside their field. By contrast Design patents have the most limited application.

5 Business Formation and Outcomes

Having established how patents associated with household innovations differ from traditional patents, this section looks at the types of business associated with household innovations; their characteristics, innovation dynamics and outcomes. The goal is to assess whether the innovator is able to monetize their innovation, either through increased business income possibly from licensing or use of the patent. There are other ways the inventor might monetize their innovation that we do not observe here such as through direct payments.[20] It should be noted that the majority of patented household innovations are not directly tied to a business that the inventor owns. Table 5 shows that only 19% of patented household innovations; those accounted for by individual assignee and unassigned patents, are associated with a business. The equivalent rate for patents with a declared business assignee is 93%.

5.1 Characteristics of Patenting Firms: Industry, Age and Size

We start by looking at the industry composition of the nonemployer firms that obtain a patent. Patenting nonemployer firms are extremely rare. Out of more than 20M nonemployer firms in a typical year, only around 5,000 firms will seek out a patent (less than 0.03%).

[20] This form of income might be observed through their income tax forms.

We limit our analysis to nonemployer firms that are born after 2000. We exclude existing nonemployers born prior to avoid left censoring in the patents we can match.[21] Figure 3 shows the industry composition of patenting nonemployer firms weighted by number of patents they own, top, and that of all nonemployer firms, bottom. Figure 3 shows, a disproportionate share of patents originate at nonemployer firms that engage in Professional Services followed by Finance & Real State and Retail. This is very different from the industry composition of nonemployer firms, which is dispersed much more evenly across industries.

Businesses associated with household innovations differ from the overall population of nonemployer businesses. We are interested in understanding whether the trigger for creating these businesses is the expectation of a patent grant and thus a means to try to capitalize on an innovation or instead the business activity predates the patent application. We explore similar patterns for firms with employees. Figure 4 graphs the distribution of firm age when the firm/inventor applies for their first patent for both employer and nonemployer firms.[22] We define firm age based on the year the business first filed income taxes.

We look at applications by patenting firms in 2010. We limit our analysis to firms up to age 10. If a firm first files income taxes after the application is filed we assign a negative age equal to the difference between application year and birth year. Figure 4 shows a significant share of businesses apply for their patent before they generate revenue. The mass of distribution is to the left of their second year of business activity. Approximately 43.6% of nonemployer firms that are granted a patent apply for the patent prior to starting their nonemployer business activity. For many other businesses the birth of the business coincides with the patent application year.

A non trivial number of patent applications, 18%, are filed three or more years after starting the business activity suggesting.

[21] Currently we can only work with patent data starting in 2000. If we were to include incumbent nonemployers in 2000 there would be no way for us to determine which ones received a patent prior to 2000.

[22] We only observe granted patents.

Compared to employer businesses household innovators are more likely to start their business at time of application, although the two distributions are centered around age zero. The tighter distribution for nonemployers can be attributed to the shortened lifecycle of nonemployer firms, most of whom are very short-lived with more than 50% of nonemployer firms exiting before year 2 and 70% of nonemployer firms exiting by year 3 (Fairlie and Miranda, 2017)

We are interested in understanding the revenue generated by household innovations vis-a-vis innovations tied to established employer businesses. Figure 5 looks at the revenue distribution for firms that own patents as a function of their employer status. As before we focus on the cross section of firms age 10 or less in 2010. Revenue follows a log-normal pattern with the distribution centered at $10,000 for household innovations.[23] Revenue for innovative employer businesses is similarly shaped but centered around much larger revenues of $1.2 million. Businesses associated with household innovations do not appear to generate much income on average at time of application. There is however a fairly wide distribution with a standard deviation of $97,500.

Figure 6 looks at income growth before and after patent grant. To avoid composition effects as a result of firm exit we show results for a balanced panel of firms that survive for a minimum of 5 years. For comparison we show revenue for employer businesses. We normalize revenue to equal 100 at grant time, t, to facilitate comparison with employer businesses. Figure 6 shows income growth prior to patent grant is considerable and very similar for both employer and nonemployer business. In the two years prior to patent grant revenue grows by 25% relative to the base. Income plateaus for nonemployer business shortly after grant and starts declining one year after. Very few firms transition to employer status so this pattern is not due to excluding successful exits out of nonemployment.

Revenue growth by employer businesses seems to be very different after grant. These

[23] It should be noted that firms who patent prior to starting their business (negative age firms) are not included in the distribution

firms display an acceleration of revenue that seems to exhaust itself two years after grant. Overall these results suggest that on average household innovators are not as successful in capitalizing their innovations after grant.

5.2 Dynamics and Transition to Employer Status

Finally, we look at the probability that a nonemployer business hires employees in particular whether patenting activity is associated with the successful growth expansion to a business that generates paid jobs for other individuals. For this exercise we focus on the cohort of new nonemployer startups in 2000 and ask ourselves how many transition into employer status each year after.

We find that of the approximately 5.24M new nonemployer entrants in 2000, around 100,000 eventually transition to employer firms over their life, for a cumulative transition rate of approximately 2%. Of this cohort, 3,700 nonemployer firms hold a patent. Of these, less than 130 will transition to employer firms over their life cycle, for a cumulative transition rate of around 3.4%, or 70% higher than non patenting firms. Annual transitions are graphed in Figure 7. As we can see, patenting firms are more than twice as likely to transition to employer firms within the first two years, relative to non patenting firms.

5.3 The value of Household Innovations

Relatively few household innovations become the foundation of a business. However, those that do give us an indication of the value of these innovations if only from the revenue they generate. Household innovations that do not directly translate into a business owned by the inventor might be expected to generate income in other ways that we do not observe in the data such as contracts or direct payments. Many other might be monetized by incumbent companies with specific market knowledge and resources to market and profit from the innovation. Many other may simply never be pursued directly but contribute to the knowledge base that generate other innovations. Other innovations might go unnoticed and yet other may simply have no value at all. Assigning value to these innovations is difficult if not impossible. However, a simple back of the envelope calculation might give us a sense of the magnitude of their overall value. To this end we calculate a range of potential values based on both the marginal and average direct income generated by businesses owned by household inventors. We focus first on innovations that are tied to nonemployer businesses. We calculate the average income generated by those business while they remain in operation. For simplicity we ignore income generated by these businesses after they hire their first employee since there are relatively few transitions. We base our calculation on the cohort of firms born in 2000 that own a patent. We track these firms through 2011 or until they exit.

Our starting point for identifying the economic value of these patents is to first come up with a revenue elasticity for each patent grant. In Table 12 we run several revenue specifications based on known factors that are seemingly unrelated to the innovation itself, but can potentially impact the revenue stream of these businesses. These include technology sector and zip code-year controls, as well as demographic controls (male, U.S. born, race and age) across the full nonemployer sample and patenters-only. In column (5), we control for selection using a Heckman selection model. The results from our specifications reveal that patents have a positive and significant impact on revenue. Across all nonemployer and patenting firms, the specifications suggest that a 10% increase in granted patents is associated with a 0.3-0.4% increase in revenue (combining the elasticities of the patent application and grant), while a 10% increase in citations is associated with a 0.03-0.06% increase in revenue. These results are consistent after controlling for selection.

In attempting to compute the economic value of these patents, we first need to tabulate the total number of household innovations as measured by patents and number of businesses associated with these patents. Table 1 and Table 2 tell us that we have approximately 93,000 matched patents to nonemployer businesses. These 93,000 patents match to 42,000 unique nonemployer businesses (2.2 patents per business). Assuming the same employer-to-nonemployer match ratios in Table 2 and applying them to the set of unmatched patents gives us 184,000 unmatched nonemployer patents, which would convert to approximately 83,000 nonemployer businesses (assuming the same ratio of patents per business). We therefore need to approximate the revenue streams for the 83,000 "missing" nonemployer businesses to tabulate the full economic impact of household innovations. Nonemployer businesses with patents generate approximately \$10,200 in annual revenue on average (versus \$9,700 generated on average for nonemployer businesses who hold no patents). Nonemployer businesses with patents also have an average survival rate of 3.95 years (versus 2.72 years for nonemployer businesses without patents). Therefore, if we take the aggregate lifetime revenue of the 42,000 nonemployer businesses with patents, we get an economic value of approximately \$1.7B (or \$18,500 per patent). Applying the same values to the 83,000 "missing" nonemployer businesses with patents gives us a cumulative economic value of \$5.0B for all household innovations between 2000 and 2011 in real 2000 dollars.

It is important to note that this calculation requires a number of strong assumptions which may differ greatly from reality. First, the revenue generated by business started by household inventors themselves is the same as the revenue generated by household innovations whose outcomes we are not able to observe including those sold to or appropriated by existing businesses. Second, businesses started by household inventors would not generate revenue where it not for the innovation. Third, the cost of developing the innovation is negligible. Finally, we have limited our analysis to patented household innovations. While arguably the most valuable innovations likely they represent but a small portion of all household innovations. We have made no effort at placing an economic value to innovations that are not known to the patent system.

6 Conclusion

Households are increasingly recognized as an important source of invention and innovation. Survey data shows independent inventors contribute substantially to consumer product innovations that are later incorporated into the products of incumbent firms. A challenge with survey data are the small sample sizes which either limits what we can learn about the most valuable innovations (the right tail of the distribution) or limits the scope of the innovations we can study. In this paper we use administrative data from the U.S. Patent and Trademark Office and the U.S. Census Bureau to describe patented household innovations in a systematic way. While patented innovations arguably represent but a very small slice of household innovations it is perhaps the most valuable one. We match these patents and their inventors to U.S. Census Bureau demographic and business data. We explore the demographic characteristics of housed inventors vis-a-vis salaried inventors, the characteristics and impact of their innovations, and their value when these inventors monetize their innovations through their own business.

We find household inventors are disproportionately born in the U.S. when compared with salaried inventors and consequently they are also relatively white. Businesses that hire inventors disproportionately hire foreign born inventors relative to their size in the population an indication these corporations might engage in brain gain by tapping foreign markets. Household inventors are disproportionately under 25 and over 55 consistent with the idea that household innovation is a leisure activity. Across the board, whether household or corporate inventors, we find a deficit in female and black inventors relative to the population as a hole.

Looking at the types of innovations we find household inventors work in technology classes disproportionately tied to consumer products such as Design, Mechanical and Other. These patents are about 1/2 as likely to be considered "radical" when compared with corporate patents. In terms of value, household innovations accumulate approximately 27-32% fewer citations on average. While their citation impact is smaller it remains remarkably high with an average of 13.6 citations per patent (through December 2016). Finally, we find relatively few household inventors start a business around their innovation. Only 19% of household innovations are tied to a business. These businesses average $10,000 in revenue at time of patent application and are more than twice as likely to transition to hire their first employee than nonemployers who do not patent.

Finally, our back-of-the-envelope calculations suggest patented household innovations granted between 2000 and 2011 may generate approximately $5.0B in 2000 dollar revenue. While this might not be extraordinary when compared to the value of corporate patents, it is non-trivial, which raises important questions about R$D and innovation policy.

To conclude, patented household innovations have impact and value. Many of them are radical and represent breakthroughs in their fields. Despite efforts to understand their role in the economy our knowledge of innovations and their inventors remains limited. Administrative data helps shed light on this population and their impact. Combined with a targeted survey of household inventors and their patented inventions could go a long way to expand our knowledge in this area.

References

Arora, A., W. M. Cohen, and J. P. Walsh (2016) "The acquisition and commercialization of invention in American manufacturing: Incidence and impact," *Research Policy*, 45(6), 1113–1128.

Bell, A., R. Chetty, X. Jaravel, N. Petkova, and J. Van Reenen (2016) "The lifecycle of inventors," .

Bell, A., et al. (2016) "The lifecycle of inventors," .

Bureau of Labor Statistics (2016) "Foreign Born Workers: Labor Force Characteristics 2015," Press Release, https://www.bls.gov/news.release/pdf/forbrn.pdf. Published 2016-05-19. Accessed 2017-02-18.

Dahlin, K. B., and D. M. Behrens (2005) "When is an invention really radical?: Defining and measuring technological radicalness," *Research Policy*, 34(5), 717–737.

de Jong, J. P. (2016) "Surveying innovation in samples of individual end consumers," *European Journal of Innovation Management*, 19(3), 406–423.

Deming, W. E. (1990) *Sample design in business research*, vol. 23. John Wiley & Sons, illustrated, reprint, revised edn.

Dreisigmeyer, D., S. Graham, C. Grim, T. Islam, A. Marco, and J. Miranda (2014) "A Patent Classification System for the Business Dynamics Statistics," Discussion paper, mimeo.

Fairlie, R. W., and J. Miranda (2017) "Taking the Leap: The Determinants of Entrepreneurs Hiring Their First Employee," *Journal of Economics and Management Strategy*, 26(1), 3–34.

Franke, N., and S. Shah (2003) "How communities support innovative activities: an exploration of assistance and sharing among end-users," *Research policy*, 32(1), 157–178.

Graham, S. J., C. Grim, T. Islam, A. C. Marco, and J. Miranda (2018) "Business dynamics of innovating firms: Linking US patents with administrative data on workers and firms," *Journal of Economics & Management Strategy*, 27(3), 372–402.

Hall, B. H., A. B. Jaffe, and M. Trajtenberg (2001) "The NBER patent citation data file: Lessons, insights and methodological tools," Discussion paper, National Bureau of Economic Research.

Hall, B. H., and M. Trajtenberg (2004) "Uncovering GPTs with patent data," Discussion paper, National Bureau of Economic Research.

Intellectual Property Owners Association (2015) "IP Spotlight 2015," Web Report, https://www.bls.gov/news.release/pdf/forbrn.pdf. Published 2015-07-31. Accessed 2017-02-18.

Jones, B. F. (2009) "The burden of knowledge and the death of the renaissance man: Is innovation getting harder?," *The Review of Economic Studies*, 76(1), 283–317.

Kim, J., and G. Marschke (2015) "Teams in R&D: Evidence from US Inventor Data," .

Lüthje, C. (2004) "Characteristics of innovating users in a consumer goods field: An empirical study of sport-related product consumers," *Technovation*, 24(9), 683–695.

Serrano, C. J. (2010) "The dynamics of the transfer and renewal of patents," *RAND Journal of Economics*, 41(4), 686–708.

Shah, S. (2000) "Sources and patterns of innovation in a consumer products field: Innovations in sporting equipment," *Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA, WP-4105*.

von Hippel, E. (1976) "The dominant role of users in the scientific instrument innovation process," *Research policy*, 5(3), 212–239.

von Hippel, E., J. P. J. de Jong, and S. Flowers (2012) "Comparing Business and Household Sector Innovation in Consumer Products: Findings from a Representative Study in the United Kingdom," *Management Science*, 58(9), 1669–1681.

Wagner, D., and M. Layne (2014) "The Person Identification Validation System (PVS): Applying the Center for Administrative Records Research and Applications (CARRA) Record Linkage Software," Discussion paper.

Table 1. U.S. Patents by Assignee Type and Year

	Individual	Business	Unassigned	Total
2000	970	79,500	21,500	107,000
2001	980	82,900	20,100	109,000
2002	930	81,200	19,000	106,000
2003	890	82,900	18,300	107,000
2004	860	80,100	16,300	101,000
2005	790	71,400	13,500	89,000
2006	980	88,700	16,200	110,000
2007	870	81,600	14,900	101,000
2008	760	81,400	14,300	99,400
2009	850	84,700	13,400	102,000
2010	960	108,000	16,500	130,000
2011	950	110,000	15,900	130,000
Total	10,800	1,032,000	200,000	1,290,000

Source: Authors calculations based on public USPTO data on granted patents by US entities between 2000-2011. Notes: Counts are rounded to comply with disclosure requirements.

Table 2. Percentage of Patents: by Assignee Type, Type of Business & Year

	Individual			Business			Unassigned		
	E	NE	U	E	NE	U	E	NE	U
2000	2.1	57.9	40	91.5	1.8	6.7	0	28.8	71.2
2001	1.4	63.8	34.7	91.6	1.9	6.5	0	29.2	70.8
2002	(D)	55.5	44.5	92	1.7	6.3	0	23.5	76.5
2003	(D)	56.9	43.1	92.4	1.7	5.9	0	23.7	76.3
2004	(D)	53.8	46.2	92.2	1.7	6.1	0	23.9	76.1
2005	1.6	55.6	42.7	91.8	1.8	6.4	0	25.3	74.7
2006	2	51.4	46.6	91.8	1.8	6.3	0	23.7	76.3
2007	1.8	52.3	45.9	92.2	1.7	6.1	0	21.5	78.5
2008	2.6	48.1	49.3	92.2	1.7	6.1	0	21.4	78.6
2009	1.4	50.5	48.1	92.3	1.7	6.1	0	21.4	78.6
2010	1.3	55.5	43.3	92	1.7	6.2	0	23.5	76.5
2011	1.7	56.3	42	90.9	1.9	7.2	0	23.9	76.1
Total	1.3	55	43.7	91.9	1.8	6.3	0	24.4	75.6

Source: Authors' calculations based on public USPTO data on granted patents applied for between 2000-2011. Notes: Type of business: E=Employer, NE=Non-Employer, U=Unknown. (D) identifies suppressed values.

Table 3. Inventor Demographics: by Assignee Type and Type of Business

	Individual			Business			Unassigned		
	E	NE	U	E	NE	U	E	NE	U
Male	86.5	91.4	90.6	92.1	91.8	90.7	(0)	89.3	87.7
U.S. Born	72.1	82.1	80.3	66.1	67.4	63.2	(0)	82.8	81.3
Black	1.8	2.1	3.5	0.9	0.9	1	(0)	3.1	4.2
White	78.4	84.8	83.1	73.9	75.8	72.9	(0)	84.2	83.1
Other	19.8	13.1	13.4	25.2	23.3	26.1	(0)	12.6	12.8
Age $<$ 25	1.8	1.6	3.9	0.5	1.2	0.9	(0)	2	2.3
25 $<$ Age $<$55	73.9	67.1	58	81	75.3	77.1	(0)	65	63.1
Age $>$55	24.3	31.4	38.1	18.5	23.5	22	(0)	33	34.5
Total Inventors*	110	6,600	1,200	1,320,000	19,200	77,100	(0)	37,300	38,400
Total Patents*	60	4,700	1,100	666,000	10,800	40,400	(0)	31,100	35,100

Source: Authors calculations based on public USPTO data on granted patents applied for between 2000-2011. Notes: Type of business: E=Employer, NE=Non-Employer, U=Unknown. * Counts are rounded to comply with disclosure requirements. (D) identifies suppressed values.

Table 4. Percent of U.S. Patents by Assignee Type and Technology Class

	Individual	Business	Unassigned
Chemical	6.9	10.7	5
C&C	11.3	29.4	5.8
Design	19.8	9.2	27.1
D&M	10	11.4	6.4
E&E	8.6	18.2	8.1
Mechanical	16	10.6	17.5
Others	26.7	10.1	29
Plant	0.6	0.4	1.1
Total*	10,800	1,030,000	200,000

Source: Authors calculations based on public USPTO data on granted patents applied for between 2000-2011. Notes: Technology Class: C&C=Computers & Communications, D&M=Drugs & Medical, E&E=Electrical & Electronic. * Total patent counts in this row are rounded to comply with disclosure requirements.

Table 5. Patent Technology Class: Percent by Assignee Type and Type of Business

	Individual Assignee			Business Assignee			Unassigned		
	E	NE	U	E	NE	U	E	NE	U
Chemical	7	8.6	4.8	10.7	10	10.8	0	6.1	4.6
C&C	16.8	15.1	6.3	29.5	29.4	28.1	0	9.3	4.6
Design	14	3	41.1	9.3	6.9	9.4	0	4.8	34.3
D&M	25.9	11.6	7.6	11	13.5	16.6	0	7.8	6
E&E	15.4	10.4	6.2	18.5	13	15.2	0	9	7.8
Mechanical	11.2	18.9	12.6	10.8	9.8	8.7	0	22.5	15.8
Others	7.7	32.3	20.3	10	13.9	9.9	0	40.4	25.3
Plant	2.1	0	1.2	0.3	0.9	0.9	0	0.1	1.4
Total*	140	5,900	4,700	949,000	18,000	65,500	0	49,000	151,000

Source: Authors calculations based on public USPTO data on granted patents applied for between 2000-2011. Notes: Each column adds up to one. Technology Class: C&C=Computers & Communications, D&M=Drugs & Medical, E&E=Electrical & Electronic. Type of business: E=Employer, NE=Non-Employer, U=Unknown. * Total patent counts in this row are rounded to comply with disclosure requirements. (D) identifies suppressed values.

Table 6. Demographic Differences by Technology Class: Non-Employer versus Employer

	Male	US Born	Black	White	Other	Age<25	25<Age<55	Age>55
Chemical	0.8*	7.1***	0.6***	5.3***	-5.9***	1.2***	-11.3***	10.2***
C&C	0.3	6.6***	0.6***	4.5***	-5***	0.5***	-9.9***	9.3***
Design	-6.1***	-0.9	4.3***	-5.8***	1.5*	1.7***	-8.7***	7***
D&M	2.1***	5***	0.7***	3.9***	-4.6***	0.7***	-11***	10.3***
E&E	0.1	8.2***	0.7***	7.4***	-8.2***	1.4***	-13.2***	11.8***
Mechanical	-0.9***	7***	0.9***	4.8***	-5.8***	1.2***	-11.6***	10.4***
Others	-5.4***	6.4***	2***	2.8***	-4.7***	1.1***	-8.9***	7.8***
Plant	1	28.8***	-0.7	-12.5**	13.2***	1.9	-12.2	10.3
Total	-1***	10.4***	1.1***	7.1***	-8.3***	1.1***	-11.8***	10.8***

Source: Authors calculations based on public USPTO data on granted patents applied for between 2000-2011. Notes: Numbers represent the difference in the proportion of patents between nonemployer and employer patents. Technology Class: C&C=Computers & Communications, D&M=Drugs & Medical, E&E=Electrical & Electronic. Type of business: E=Employer, NE=Non-Employer, U=Unknown. * $p<0.05$, ** $p<0.01$, *** $p<0.001$

Table 7. Mean Team Size: by Technology Class, Assignee Type and Type of Business

	Individual Assignee			Business Assignee			Unassigned		
	E	NE	U	E	NE	U	E	NE	U
Chemical	3.1	1.93	1.38	3.06	2.47	2.86	0	1.48	1.32
C&C	2.67	1.88	1.31	2.65	2.32	2.53	0	1.47	1.27
Design	1.9	1.68	1.42	2.21	1.7	2.11	0	1.28	1.19
D&M	2.92	1.9	1.46	3.1	2.45	2.91	0	1.53	1.39
E&E	2.18	1.76	1.31	2.56	2.13	2.4	0	1.38	1.22
Mechanical	1.63	1.68	1.24	2.49	1.98	2.3	0	1.29	1.15
Others	1.73	1.67	1.27	2.44	1.98	2.28	0	1.29	1.15
Plant	2	1	1.04	1.25	1.15	1.3	0	1.68	1.31
All Patents	2.38	1.8	1.36	2.65	2.24	2.49	0	1.37	1.21

Source: Authors calculations based on public USPTO data on granted patents applied for between 2000-2011. Notes: Technology Class: C&C=Computers & Communications, D&M=Drugs & Medical, E&E=Electrical & Electronic. Type of business: E=Employer, NE=Non-Employer, U=Unknown. (D) identifies suppressed values

Table 8. Mean Citation Count: by Technology Class, Assignee Type and Type of Business

	Individual Assignee			Business Assignee			Unassigned		
	E	NE	U	E	NE	U	E	NE	U
Chemical	8.9	7.52	7.25	10.42	8.78	11.77	0	7.08	6.22
C&C	17.04	23.21	19.89	20.34	23.92	21.93	0	16.41	14.41
Design	6.65	8.58	6.18	12.32	11.38	10.7	0	8.55	7.07
D&M	18.86	23.73	17.65	25.73	22.81	21.75	0	18.66	15.32
E&E	12.09	13.1	9.61	13.84	16.26	15.14	0	10.6	8.88
Mechanical	10.88	7.81	7.78	11.79	14.88	12.64	0	8.13	6.56
Others	9	9.98	8.98	14.45	12.39	12.3	0	8.03	7.72
Plant	1.67	0.5	0.36	0.31	0.35	0.32	0	0.76	0.29
All Patents	13.1	13.28	9.34	16.36	16.49	16.81	0	10.13	8.09

Source: Authors calculations based on public USPTO data on granted patents applied for between 2000-2011. Notes: We exclude patents with zero citations. Technology Class: C&C=Computers & Communications, D&M=Drugs & Medical, E&E=Electrical & Electronic. Type of business: E=Employer, NE=Non-Employer, U=Unknown. (D) identifies suppressed values

Table 9. Pseudo-Maximum Log Likelihood Regression on Patent Citations

	(1)	(2)
Dependent Variable	Citations	Citations
Grant Year	-0.1616***	-0.1622***
	(0.00760)	(0.00757)
Team Size	0.07265***	0.06831***
	(0.00426)	(0.00403)
Employer Patents	0.23618***	
	(0.02269)	
Non-Employer Patents	0.07342***	
	(0.01385)	
Unmatched Patents	Dropped	Dropped
Firm Assigned Patents		0.04132
		(0.02733)
Individual Assigned Patents		-0.2460***
		(0.02746)
Unassigned Patents	Dropped	Dropped
USPC Fixed Effects	Yes	Yes
Constant	326.192***	327.522***
	(14.7814)	(14.7207)
Observations	1,290,000	1,290,000

Standard Errors are clustered at the USPC Technology Class level. * $p<0.05$, ** $p<0.01$, *** $p<0.001$

Table 10. Proportion of Radical Patents (per thousand): by Technology Class, Assignee Type and Type of Business

	Individual Assignee			Business Assignee			Unassigned		
	E	NE	U	E	NE	U	E	NE	U
Chemical	0	3.9	4.5	18.1	17.5	17.6	0	1.7	2
C&C	41.7	14.5	6.7	14.1	18.7	15.5	0	5.7	3.7
Design	0	11.2	9.8	28.4	19.8	21.9	0	9.4	12.3
D&M	0	13.1	8.4	25.6	22.8	22.3	0	3.4	3.2
E&E	0	6.5	0	13.2	15.2	14.8	0	2.3	3.1
Mechanical	0	4.5	3.4	12	17.8	16.5	0	2.8	2.3
Others	0	5.7	3.1	15.2	13	16.5	0	1.8	1.4
Plant	0	0	0	1.6	0	5.2	0	0	0.5
Total	7	7.8	6.4	16.8	17.2	17.4	0	2.9	5.7

Source: Authors calculations based on public USPTO data on granted patents applied for between 2000-2011. Notes: Technology Class: C&C=Computers & Communications, D&D=Drugs & Medical, E&E=Electrical & Electronic. Type of business: E=Employer, NE=Non-Employer, U=Unknown

Table 11. Mean (Modified) Generality Index: by Technology Class, Assignee Type and Type of Business

	Individual Assignee			Business Assignee			Unassigned		
	E	NE	U	E	NE	U	E	NE	U
Chemical	0.6	0.64	0.62	0.6	0.61	0.59	0	0.63	0.65
C&C	0.58	0.59	0.6	0.63	0.62	0.63	0	0.62	0.63
Design	0.87	0.8	0.86	0.88	0.84	0.88	0	0.79	0.86
D&M	0.63	0.68	0.71	0.66	0.66	0.65	0	0.68	0.71
E&E	0.62	0.66	0.69	0.66	0.64	0.65	0	0.66	0.68
Mechanical	0.61	0.72	0.71	0.67	0.66	0.66	0	0.7	0.72
Others	0.64	0.69	0.7	0.67	0.68	0.67	0	0.69	0.7
Plant	1	1	1	0.99	1	0.99	0	1	1

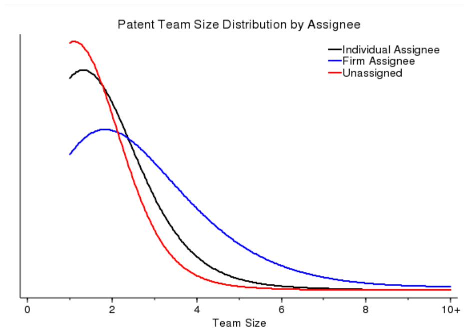
Source: Authors calculations based on public USPTO data on granted patents applied for between 2000-2011. Notes: Technology Class: C&C=Computers & Communications, D&D=Drugs & Medical, E&E=Electrical & Electronic. Type of business: E=Employer, NE=Non-Employer, U=Unknown

Table 12. Regression Results on Non Employer Revenues

	(1)	(2)	(3)	(4)	(5)
Regression	OLS	OLS	OLS	OLS	Heckman Selection (Second Stage)
Sample	All	Patenters Only	All	Patenters Only	All
Citations	0.0048***	0.0060***	0.0045***	0.0056***	0.0039***
	(0.0003)	(0.0009)	(0.0003)	(0.0009)	(0.0008)
Patent Applications	0.0305***	0.0260***	0.0262***	0.0175***	0.0254***
	(0.0009)	(0.0039)	(0.009)	(0.0039)	(0.0011)
Patent Grants	0.0135***	0.0078***	0.0117***	0.0079***	0.0139***
	(0.0010)	(0.0023)	(0.0009)	(0.0023)	(0.0011)
Team Size	-0.0019***	-0.0027***	-0.0027***	-0.0035***	-0.0026***
	(0.0003)	(0.0009)	(0.0003)	(0.0009)	(0.0004)
Demographic Controls	No	No	Yes	Yes	Yes
Zip-Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Patent-Sector Fixed Effects	Yes	Yes	Yes	Yes	Yes
R-Squared	0.019	0.259	0.062	0.278	
Observations	198,110,000	41,500	198,110,000	41,500	198,110,000

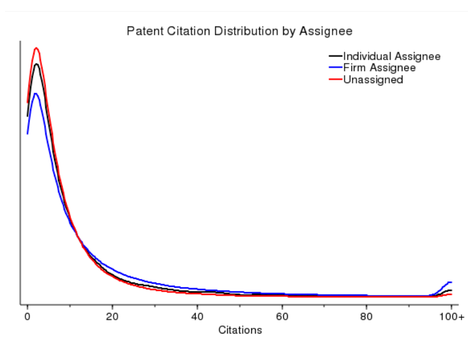
Robust Standard Errors are clustered at the patent-sector level. Selection equation for Column 5 includes demographic controls and Zip-Year fixed effects. The selection coefficient is -6.0557 with s.e. 0.0628 and is significant to the 0.1%. * p<0.05, ** p<0.01, *** p<0.001

Figure 1. Kernel Distribution of Team Size by Assignee Type, 2000-2011



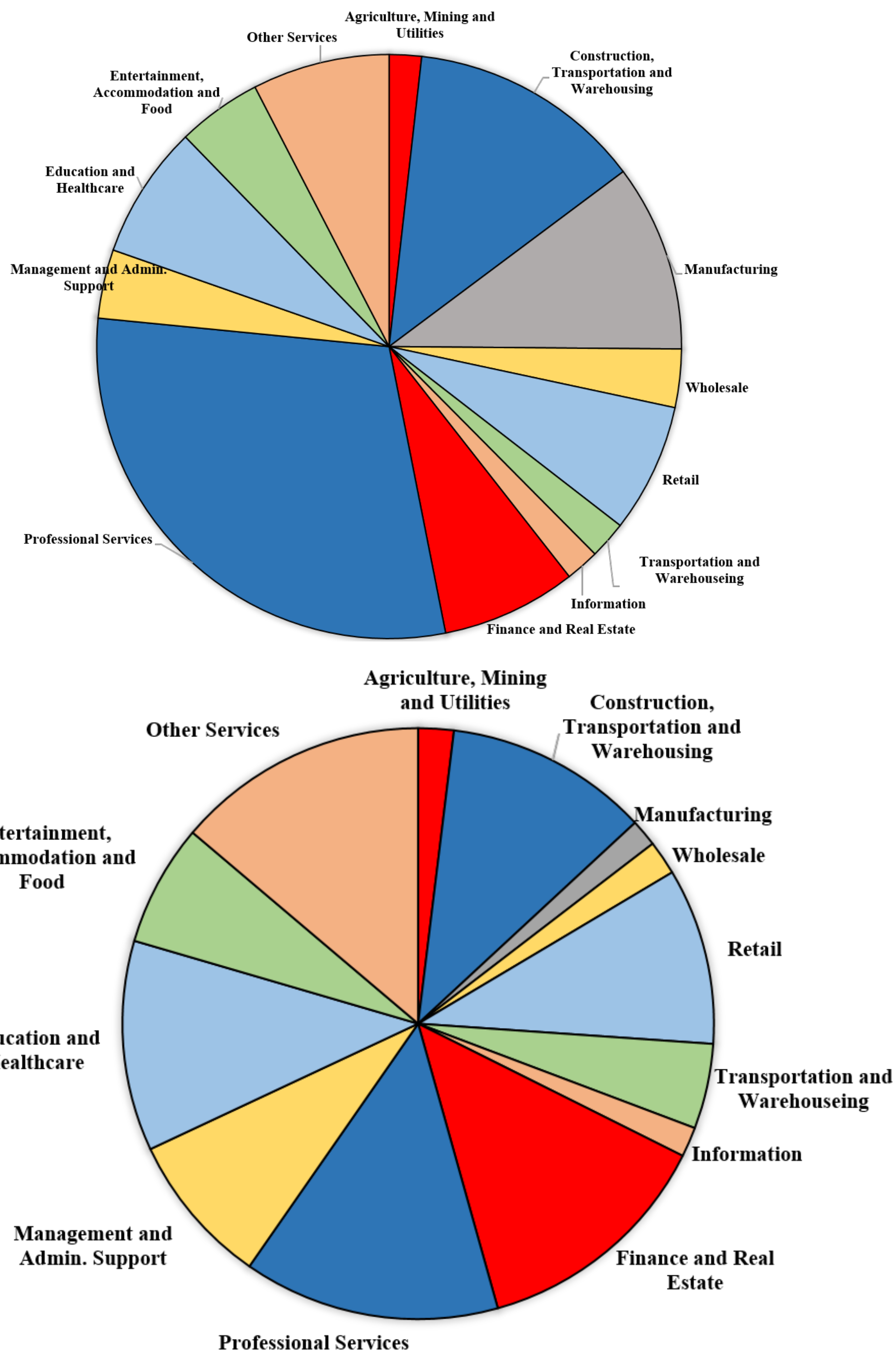
Source: Own calculations based on USPTO data on granted patents applied for between 2000-2011.

Figure 2. Kernel Distribution of Citation Counts by Assignee Type, 2000-2011



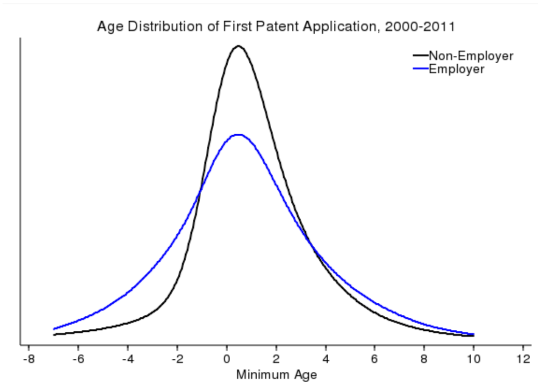
Source: Own calculations based on USPTO data on granted patents applied for between 2000-2011.

Figure 3. Industry Composition of Nonemployer Firms: Patenting (top)/All (bottom), 2000-2011



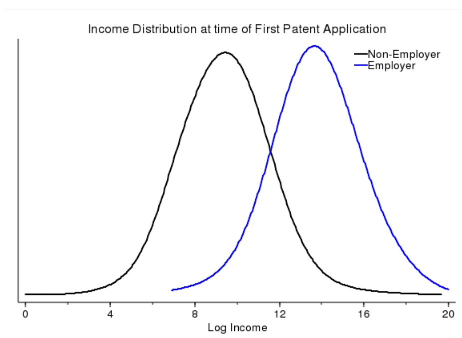
Source: Own calculations based on USPTO and U.S. Census Bureau data. Granted patents applied for between 2000-2011.

Figure 4. Kernel Distribution of Age of Nonemployer Firm for First Patent, 2010



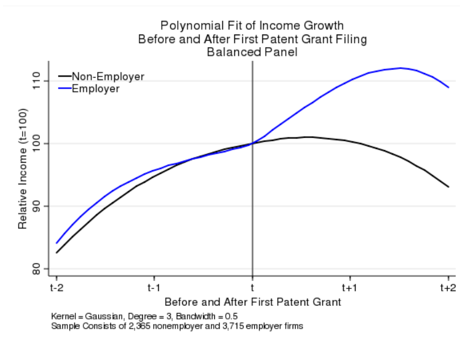
Source: Own calculations based on USPTO and U.S. Census Bureau data on patent holding firms age 10 years or less in 2010.

Figure 5. Kernel Distribution of Size of Nonemployer Firm for First Patent, 2000-2011



Source: Own calculations based on USPTO and U.S. Census Bureau data on granted patents applied for between 2000-2011.

Figure 6. Total Income Before and After First Patent, Balanced Panel



Source: Own calculations based on USPTO and U.S. Census Bureau data on granted patents applied for between 2000-2011. Sample includes a balanced panel of patenting firms centered at patent grant.

Figure 7. Transition to Employer Firms by Year, 2000 Cohort

Source: Own calculations based on USPTO and U.S. Census Bureau data. 2000 cohort of nonemployer business.

Appendix A Matching Process and Data Construction

In this section, we outline the matching process between USPTO granted patents and the full non-employer dataset (iLBD) at Census. We start by describing the individual datasets and features of the datasets that will be matched. We then outline the matching algorithm and post a number of statistics on the match rates across different patent types.

A.1 USPTO Patent Data

The USPTO patent database consists of all granted patents applied for between 2000 and 2011 by US entities and excludes all foreign entities, as well as government patents.[24] Counts of domestic patents with inventor and assignee data are plotted in Figure 1.

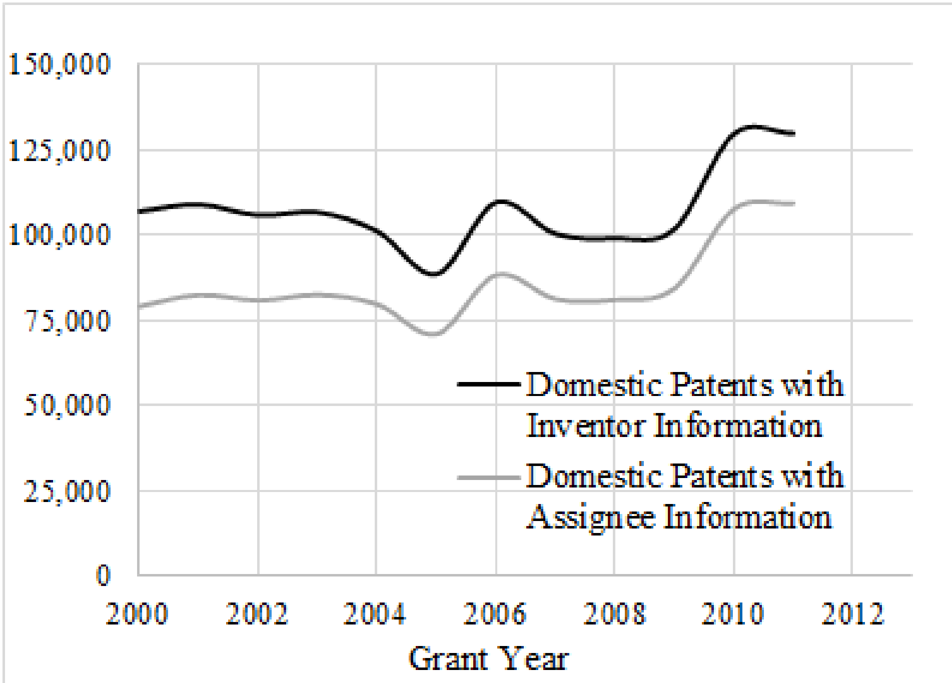
Our matching algorithm attempts to create name and address matches from two distinct sources of information contained in USPTO patent record: i) The assignee, typically a firm, for whom patent ownership belongs, and ii) the inventor – persons who may or may not be affiliated with a firm. In cases where no assignee is named, it is assumed that the patent's ownership remains with the inventor(s).

We compile our matching database from two distinct sources of PatentsView data from USPTO, each associated with either the Assignee or the Inventor.

Our matching algorithm attempts to create name and address matches from two distinct sources of information contained in USPTO patents: i) The assignee, typically a firm, for whom patent ownership belongs, and ii) the inventor persons who may or may not be affiliated with a firm that came up with the patent. In cases where no assignee is name, it is assumed that the patents ownership belongs to the inventor and/or inventors. We are primarily interested in collecting names and any geographical data associated with the

[24] We only keep patents of assignee type "02 - US Company and/or Corporation" and type "04 - US Individual", as well as patents with missing assignee information that originate in the US and contain US inventor data.

Figure A1. Mean (Modified) Annual Patent Application Counts of Granted US Patents by Application Year, 2000-2013



patents. We compile our matching database from two distinct sources of data from USPTO, each associated with either the Assignee or the Inventor.

A.1.1 Cleaning of USPTO Assignee Data

The matching information for assignees is limited to the firm name, city and state. We use city and state as our blocking variables and allow for fuzzy matching based on name. We start with approximately 1.29M patent observations across all years and drop around 260,000 patents that do not have an assignee name to match against, leaving us with 1.03M patents to match assignee information against. Nearly all of the 1.03M patents have geographic information, including city and state to match against.

In each year, there are on average 18,000 unique assignee names to match against and slightly more geographic pairs, indicating that a small subset of assignees apply for patents from multiple locations. The total number of unique assignees between 2000 and 2011 is

Table A1. Assignee Counts from USPTO data on granted patents by US entities, 2000-2011

	All Patents	Domestic Patents with Assignees	Unique Assignee-Geographic Pairs	Unique Assignees
2000	107,000	79,500	20,800	18,800
2001	109,000	82,900	21,000	18,900
2002	106,000	81,200	19,600	17,800
2003	107,000	82,900	19,200	17,700
2004	101,000	80,100	18,600	17,200
2005	89,000	71,400	17,100	15,900
2006	110,000	88,700	19,900	18,300
2007	101,000	81,600	18,300	17,000
2008	99,400	81,400	17,900	16,700
2009	102,000	84,700	17,900	16,700
2010	130,000	108,000	21,200	19,800
2011	130,000	110,000	21,700	20,200
Total	1,290,000	1,030,000	123,000	102,000

Source: Authors calculations on public USPTO data on granted patents applied for by US entities between 2000-2011. Notes: Counts are rounded to comply with disclosure requirements.

approximately 102,000 and provide potential matches for 1.03M patents (80% of possible matches).

A.1.2 Cleaning of USPTO Inventor Data

Inventors are listed separately from the Assignees and are considered wholly different as they are typically employees of the assignee firms. Inventor data contains a separate identifier for each inventor and also contains city and state level geographic data. Multiple inventors can work on each patent. The number of inventors greatly exceeds the number of assignees. Because the Integrated Longitudinal Business Database (iLBD) mainly consists of person-level identifiers, inventors will serve as a primary matching criteria.

In each year, there are around 160,000 unique inventor names on average to match the iLBD against and nearly 1M unique individuals associated with patents granted between 2000-2011. Nearly all of the data contains geographic information of some form, including city and/or state, with a small proportion of inventors applying for patents across multiple locations. Combining this data with the Assignee data gives us the full matching criteria

Table A2. Inventor Counts from USPTO data on granted patents by US entities, 2000-2011

	All Patents	Domestic Patents with Inventors	Unique Inventor-Geographic Pairs	Unique Inventors
2000	107,000	107,000	161,000	153,000
2001	109,000	109,000	166,000	158,000
2002	106,000	106,000	165,000	157,000
2003	107,000	107,000	169,000	160,000
2004	101,000	101,000	164,000	156,000
2005	89,000	88,900	149,000	142,000
2006	110,000	110,000	177,000	167,000
2007	101,000	101,000	166,000	157,000
2008	99,400	99,300	166,000	157,000
2009	102,000	102,000	175,000	165,000
2010	130,000	130,000	220,000	205,000
2011	130,000	130,000	222,000	207,000
Total	1,290,000	1,290,000	1,200,000	990,000

Source: Authors calculations on public USPTO data on granted patents by US entities between 2000-2011
Notes: Counts are rounded to comply with disclosure requirements.

to perform our name and address match. To summarize our matching frame, we have approximately 180,000 unique inventors and assignees to match the iLBD against in every year. These 180,000 inventors and assignees represent around 110,000 patents in each year for 1.29M total patents.

A.2 Integrated Longitudinal Business Database Cleanup

On the non-employer side of the data, we start by combining all of the individual cross-sections of the Integrated Longitudinal Business Database (iLBD) starting in 2000 until 2011. The iLBD consists of both non-employer businesses (identified with an Employer Identification Number, EIN) and sole-proprietorships (identified by a Protected Identification Key, PIK). The breakdown and counts of businesses of each type are as follows:

The identifying information used to link to the patents consists of a name, city and state, along with a unique identifier that is able to link non-employer businesses over time. Names are given by two separate name variables. We separate the two name variables and

Table A3. Non-Employer Businesses Counts by Type

Year	Non-Employer Businesses	Non-Employer EIN	Non-Employer PIK
2000	16,530,000	2,120,000	14,410,000
2001	16,980,000	2,230,000	14,750,000
2002	17,650,000	2,270,000	15,380,000
2003	18,650,000	2,420,000	16,230,000
2004	19,520,000	2,530,000	16,990,000
2005	20,390,000	2,670,000	17,720,000
2006	20,770,000	2,590,000	18,180,000
2007	21,710,000	2,620,000	19,090,000
2008	21,350,000	2,540,000	18,810,000
2009	21,700,000	3,000,000	18,700,000
2010	22,110,000	3,000,000	19,110,000
2011	22,490,000	3,050,000	19,440,000
Total	239,850,000	31,040,000	208,810,000

Source: Nonemployer Statistics

Notes: Counts are rounded to comply with disclosure requirements.

stack them with their unique identifier in order to obtain every name combination in the database. In addition, approximately 55% of the names consist of two individuals separated by an ampersand, such as "John & Jane Doe". We separate out each of these observations into two observations (e.g. "John Doe" and "Jane Doe"). All together, these combinations give us more than 297M+ unique observations for the 183M non-employer businesses to match against.

A.3 Matching Algorithm

Once the two matching datasets have been completed, we run the following name and address matching algorithm in order of best possible match to worst possible match.

a. Name, City and State Only the inventor dataset of the USPTO contains CITY data

b. Name and State Includes both Inventor and Assignee data and consists of the largest possible match

c. Name Only Worst possible match set, but we can keep unique matches

Table A4. Number of Patent Matches by Match Criteria, 2000-2011

	Number of Matches	% of Total Matches
Match Criteria 1 - Name, City and State	713,000	69%
Match Criteria 2 - Name and State	207,000	20%
Match Criteria 3 - Name Only	117,000	11%
Total	1,037,000	

Source: Authors calculations using iLBD Data. Notes: Counts are rounded to comply with disclosure requirements.

Table A5. Breakdown of Matches by Identifier, 2000-2011

	Matched Patents	Inventor Only	Assignee Only	Both
Match Criteria 1 - Name, City and State	713,000	500,000	102,000	112,000
Match Criteria 2 - Name and State	207,000	130,000	53,000	24,000
Match Criteria 3 - Name Only	117,000	78,000	26,000	13,000
Total Patents	1,037,000	708,000	181,000	149,000

Source: Authors calculations using iLBD Data. Notes: Counts are rounded to comply with disclosure requirements.

We use SAS PROC DQMATCH algorithm to run the match. After each step, we only keep the residual non-matched patents so that each patent can only be matched according to one of the criteria sets above. Table A4 provides summary statistics on the full match rates by step. These consist of the raw matches (prior to any cleaning).

We are able to match approximately 80% of the 1.29M patents that we start out with. More than two-thirds of the matches occur at the highest quality where the patent's assignee/inventor's name, city and state matched a non-employer business' name, city and state. Approximately 1/5 of the matches occur at the "name and state" resolution, with the remaining matches occurring at the "name" resolution. Each of these matches can occur through an inventor match, assignee match, or for some patents, in both. The breakdown of match by identifier is:

Nearly 70% of the matches occur through the inventor, which is expected since nearly 90% of the patent matching criteria are through the inventor. About 14% of patents are matched through both the inventor and assignee, with the remaining being matched through the assignee. The next step in the matching process is to filter out the patents that are actually linked with employer firms, keep patents that have identified inventors in the PVS process, drop duplicate matches (e.g. more than one identifier for a patent-name combination) and finally augment our data using unique PVS'ed patents.

A.4 Cleaning the Set of Matches

Starting with our set of 103M matches, the first step in the cleaning process is to remove all of the patents associated with employer firms using an existing Census firm-level crosswalk (see Graham et al. 2015). These patents may have matched to the non-employer data either through the inventor who is employed by an employer firm who is the assignee, or if the name of the non-employer business is very similar or identical to the name of an employer businesses. The existing Census firm-level crosswalk exists from 2000-2011 and covers more than 1.5M patents, of which 958,000 originate in the US, with the remaining belonging to foreign assignees with US subsidiaries. This crosswalk was created using a triangulation of name-address matching of assignee data merged with linked employee-employer inventor data. The crosswalk covers around 90% of all domestic patents with firm assignees. Filtering out the employer patents will remove approximately 80% of the patents matched to the non-employer data (838,000 patents were removed). This is suggestive that a large percentage of inventors at employer firms also have non-employer businesses. Not all of the patents from these inventors are removed from the final dataset, rather only the patents that are identified to being assigned to an employer firm.

The next step in the cleaning of the matches involves filtering out the matches that have not been linked to Census data using the Census Bureau's Person Identification Validation System (PVS). The PVS process assigns an anonymous, unique person identifier (PIK) to individuals using name and address information and matching it against the Social Security Administration's numerical identification file ("Numident"). The matching process is probabilistic and it is possible for an individual to receive multiple identifiers (PIKs), especially if the provided only partial information. The USPTO patent data underwent the full PVS process for the original Census firm-level crosswalk, generating PIKs for all of the inventors identified in patents, based on names and a zip code. Because the information used to generate these matches is rather coarse (only name and zip), approximately 30% of the patent-inventor combinations have a unique identifier (PIK), while 75% have fewer than 5 identifiers. The zip code is the unique characteristic here that we miss in our non-employer matching process and hence, can be used to validate our existing matches. Our filter involves directly linking all of the PIKs assigned to each patent from the PVS process and merging them with the PIKs generated in the non-employer matches. We drop patents that were matched to the non-employer through the inventor name but are not identified in the PVS process. This removes nearly 40% of the existing matches.

The third step in the filter process drops duplicate matches by patent identifier and name. These are patents that cannot be assigned to a specific person or business because of multiple matches. There are several instances where patents match to multiple non-employer identifiers after the name and address match and after the filters have been applied. Since there is no way to distinguish between these non-employer matches, we elect to drop them. This removes 45% of existing matches.

The next step in the filter process involves "winsorizing" our existing matches by the assignee code. In this case, we count the number of patents by assignee code-year and drop the patents for the assignee code-year combinations that are in the top 0.5%. This number ranges between 20-50 patents per year. Our assumption lies in that due to size constraints, the number of patents a non-employer business can produce in a year are limited and that these observations are likely to have been missed by the existing Census firm-level crosswalk or are "unique" for entirely different reasons. This removes a further 7.5% of matches.

Finally, we augment our matches using the unique inventor identifiers from the PVS process. As mentioned earlier, approximately 30% of the patent-inventor combinations have a unique identifier (PIK). We keep the ones with the unique identifier and merge them with the full non-employer database to identify non-employer businesses that our matching methodology may have missed. We then augment our existing matches with this database.

Table A6. Filtering Out Employer Patents, 2000-2011

Grant Year	Original Match	Removal of Employer Patents	Keep PVS	Drop Duplicate	Winsorize and Augment with PVS
2000	83,800	19,700	14,400	10,500	10,200
2001	86,000	19,000	14,100	10,500	10,100
2002	84,100	18,000	11,300	8,700	8,300
2003	85,100	17,300	10,900	8,500	8,200
2004	80,900	15,900	9,900	7,600	7,400
2005	71,800	13,700	8,500	6,800	6,500
2006	88,500	16,500	9,900	8,000	7,700
2007	81,300	14,700	8,500	6,800	6,500
2008	80,600	14,300	8,000	6,400	6,100
2009	83,200	14,000	8,000	6,400	6,000
2010	106,000	18,200	10,700	8,700	8,200
2011	106,000	18,300	10,400	8,300	7,900
Total	1,037,000	200,000	125,000	97,000	93,000

Source: Authors calculations using LBD Data. Notes: Counts are rounded to comply with disclosure requirements.

This increases the number of matches by approximately 5% for a total of 68,400 matched patents. The table below summarizes the full effect of each matching stage.

This completes the matching process for the nonemployer data. Starting from 1.29M patents, we are able to successfully match 68,400 patents to the nonemployer data. The full breakdown of matches by dataset is below.

We denote the "unmatched" as unknown since a fairly large proportion of these patents were initially matched to the nonemployer dataset, but were dropped either because the inventor's personal identifier was not listed in the PVS process, or because the invention-name combination had more than one individual listed (dropped out during deduplication process). A breakdown of the "Unknown" matches is given in Table A8.

Table A8 tells us that approximately 141,000 of the 273,000 unknown patents were unmatched across all Census datasets, which implies that around 132,000 patents were linked to the nonemployer. Of these, approximately, 67% were dropped as they were not listed in the PVS process, with the remainder dropping due to either being duplicates or "winsorized".

Table A7. Total Matches by Type, 2000-2011

Grant Year	Total	Employer	Nonemployer	Unknown
2000	107,000	72,700	10,200	24,400
2001	109,000	75,900	10,100	23,200
2002	106,000	74,700	8,300	23,000
2003	107,000	76,600	8,200	22,100
2004	101,000	73,800	7,400	20,200
2005	88,900	65,500	6,500	16,900
2006	110,000	81,500	7,700	20,600
2007	101,000	75,300	6,500	18,800
2008	99,300	75,100	6,100	18,200
2009	102,000	78,200	6,000	17,800
2010	130,000	99,500	8,200	22,000
2011	130,000	99,900	7,900	22,100
Total	1,290,000	949,000	93,000	249,000

Source: Authors calculations using LBD Data. Notes: Counts are rounded to comply with disclosure requirements.

Table A8. Breakdown of Unknown Matches, 2000-2011

Grant Year	Total Unknown	Unmatched	Matched	Drop in PVS Process	Duplicates/Winsorized
2000	24,400	14,800	9,600	6,400	3,200
2001	23,200	14,200	9,000	6,000	3,000
2002	23,000	13,300	9,700	7,700	2,000
2003	22,100	13,000	9,100	7,400	1,700
2004	20,200	11,600	8,600	7,000	1,500
2005	16,900	9,600	7,300	6,100	1,100
2006	20,600	11,700	9,000	7,900	1,100
2007	18,800	10,500	8,300	7,400	900
2008	18,200	10,000	8,300	7,400	800
2009	17,800	9,700	8,100	7,200	900
2010	22,000	11,900	10,200	8,900	1,200
2011	22,100	11,600	10,500	9,500	1,100
Total	249,000	142,000	108,000	89,000	18,600

Source: Authors calculations using LBD Data. Notes: Counts are rounded to comply with disclosure requirements.

Table A9. Breakdown of PVS Process for Inventors, 2000-2011

Grant Year	Patents	PVS Patents	Inventor Names	Inventor PIKs	Inventor PIKs (Highest PVS)	Unique Inventor PIKs	Patents with Unique Inventor PIKs
2000	107,000	82,700	165,000	748,000	418,000	128,000	71,800
2001	109,000	88,400	183,000	802,000	468,000	143,000	77,600
2002	106,000	79,100	172,000	760,000	442,000	136,000	70,100
2003	107,000	80,900	180,000	787,000	470,000	143,000	72,200
2004	101,000	77,900	175,000	754,000	453,000	139,000	69,600
2005	89,000	69,700	159,000	693,000	422,000	126,000	62,500
2006	110,000	83,800	196,000	853,000	531,000	155,000	75,300
2007	101,000	74,300	177,000	773,000	496,000	139,000	66,900
2008	99,400	72,700	176,000	750,000	486,000	138,000	65,500
2009	102,000	77,000	190,000	833,000	546,000	149,000	69,500
2010	130,000	101,000	252,000	1,110,000	731,000	198,000	91,500
2011	130,000	102,000	255,000	1,130,000	755,000	199,000	91,800
Total	1,290,000	989,000	2,280,000	9,980,000	6,220,000	1,790,000	884,000

Source: Authors calculations. Notes: Counts are rounded to comply with disclosure requirements.

Appendix B Matching Demographics to Patent Data

Matching the patent data to the demographic data is a relatively straightforward process of merging multiple files and dropping duplicate matches allocated in the PVS process. We start with the original patents that have undergone the PVS process. Of our starting point of 1.29M patents, 989,000 have undergone the PVS process (76.7%). These 989,000 PVS'ed patents have 2.28M inventor names associated with the patents (average team size of approximately 2.3) and 9.98M inventor PIKs associated with them, indicating that each inventor name has on average around 4 PIKs. We start by keeping the PIK with the highest PVS score by patent-inventor combination. This removes 3.76M of the 9.98M starting inventor PIKs. We want to unduplicate the remainder of these PIKs and only keep the inventors with a unique PIK. Removing all of the duplicate PIKs associated with each inventor name leaves us with 1.79M unique inventor PIKs associated with nearly 884,000 patents from the 989,000 patents that underwent the PVS process. A yearly breakout of the counts is below.

If we break out the counts by assignee type, we find differences in the ratio of the patents that undergo the PVS process by assignee type, along with differences in the ratio of inventors that unique PIKs by assignee type. Firm assignees are most likely to have undergone the PVS process (82%), followed by individual assignees (75%), while fewer than 50% of unassigned patents undergo the PVS process. Looking at the proportion of inventors that have unique PIKs by assignee type, we find that nearly 91% of inventors in firm assigned patents have a unique PIK associated with their name. This is higher than the ratio found in individual assigned patents (83%) and the ratio in unassigned patents (76.7%). The full breakdown by assignee type is below.

Starting from the nearly 884,000 patents with unique inventor PIKs, we then merge it to the Census Numident file which contains the demographic information we are interested in. The Numident match rate is around 100%, thus completing the full demographic matching process for each patent. Turning back to the unmatched patents, we break down the match rate by sector. We show that the patents without unique PIKs and no demographic data are mainly concentrated in the "Design" and "Plant" patent sector as shown in the table below.

Appendix C Technology Classes

Table A10. Breakdown of PVS Process for Inventors by Assignee Type, 2000-2011

	Individual Assignee			Business Assignee			Unassigned		
Grant Year	Patents	PVS Patents	Patents with Unique Inventor	Patents	PVS Patents	Patents with Unique Inventor	Patents	PVS Patents	Patents with Unique Inventor
2000	970	810	650	79,500	65,100	58,300	21,500	13,400	10,200
2001	980	870	710	82,900	71,500	64,500	20,100	12,500	9,500
2002	930	660	560	81,200	66,200	60,100	19,000	8,900	6,700
2003	890	670	560	82,900	68,400	62,400	18,300	8,500	6,500
2004	860	640	550	80,100	66,600	60,700	16,300	7,700	5,900
2005	790	600	490	71,400	60,000	54,800	13,500	6,800	5,200
2006	980	700	600	88,700	72,800	66,500	16,200	7,500	5,900
2007	870	620	510	81,600	65,000	59,600	14,900	6,400	4,900
2008	760	490	430	81,400	64,100	58,700	14,300	6,000	4,600
2009	850	590	470	84,700	68,500	62,800	13,400	5,800	4,400
2010	960	720	590	108,000	89,700	82,400	16,500	7,800	6,000
2011	950	730	610	110,000	90,800	83,300	15,900	7,800	6,000
Total	10,790	8,090	6,720	1,032,000	849,000	774,000	200,000	98,900	75,800

Source: Authors calculations. Notes: Counts are rounded to comply with disclosure requirements.

Table A11. Breakdown of Demographic Match Rate by Sector, 2000-2011

Sector	Individual Assignee	Firm Assignee	Unassigned
Chemical	75.1	82.2	47.1
C&C	73.9	81	52.1
Design	11	11.4	9
D&M	75	83	50.7
E&E	75.4	82.2	43.6
Mechanical	75.6	82.1	47.4
Others	75.7	80.9	51.8
Plant	11.9	10.1	5
Total Proportion	62.1	75	38

Source: Authors calculations using LBD Data. Notes: Counts are rounded to comply with disclosure requirements.

Table A12. Technological Categories as Defined in Hall et al.(2001) Plus additions in Bold

Cat. Code	Category Name	Sub-Cat. Code	Sub-Category Name	Patent Classes
1	Chemical	11	Agriculture, Food, Textiles	8, 19, 71, 127, 442, 504
		12	Coating	106.118, 401, 427
		13	Gas	48, 55, 95, 96
		14	Organic Compounds	532, 534, 536, 540, 544, 546, 548, 549, 552, 554, 556, 558, 560, 562, 564, 568, 570, 987
		15	Resins	520, 521, 522, 523, 524, 525, 526, 527, 528, 530
		19	Miscellaneous-chemical	23, 34, 44, 102, 117, 149, 156, 159, 162, 196, 201, 202, 203, 204, 205, 208, 210, 216, 222, 252, 260, 261,349, 366, 416, 422, 423, 430, 436, 494, 501, 502, 506, 510, 512, 516, 518, 585, 588
2	Computers & Communications	21	Communications	178, 333, 340, 342, 343, 358, 367, 370, 375, 379, 385, 398, 455, 725
		22	Computer Hardware & Software	341, 380, 382, 395, 700, 701, 702, 704, 705, 706, 707, 708, 709, 710, 712, 713, 714, 902
		23	Computer Peripherals	345, 347, 726
		24	Information Storage	360, 365, 369, 711, 720, G9B
		25	Data Processing	715, 717, 718, 719
3	Drugs & Medical	31	Drugs	424, 514
		32	Surgery & Medical Instruments	128, 600, 601, 602, 604, 606, 607
		33	Biotechnology	435, 800, 930
		39	Miscellaneous - Drug & Med.	351, 433, 623
4	Electrical & Electronic	41	Electrical Devices	174, 200, 327, 329, 330, 331, 332, 334, 335, 336, 337, 338, 392, 439
		42	Electrical Lighting	313, 314, 315, 362, 372, 445
		43	Measuring & Testing	73, 324, 356, 374, 850
		44	Nuclear & X-rays	250, 376, 378, 976
		45	Power Systems	60, 136, 290, 310, 318, 320, 322, 323, 361, 363, 388, 429
		46	Semiconductor Devices	257, 326, 438, 505
		49	Miscellaneous - Elec.	191, 218, 219, 307, 346, 348, 377, 381, 386, 703, 716
5	Mechanical	51	Materials Processing & Handling	65, 82, 83, 125, 141, 142, 144, 173, 209, 221, 225, 226, 234, 241, 242, 264, 271, 407, 408, 409, 414, 425, 451, 493
		52	Metal Working	29, 72, 75, 76, 140, 147, 148, 163, 164, 228, 266, 270, 413, 419, 420
		53	Motors, Engines & Parts	91, 92, 123, 185, 188, 192, 251, 303, 415, 417, 418, 464, 474, 475, 476, 477
		54	Optics	352, 353, 355, 359, 396, 399
		55	Transportation	104, 105, 114, 152, 180, 187, 213, 238, 244, 246, 258, 280, 293, 295, 296, 298, 301, 305, 410, 440
		59	Miscellaneous - Mechanical	7, 16, 42, 49, 51, 74, 81, 86, 89, 100, 124, 157, 184, 193, 194, 198, 212, 227, 235, 239, 254, 267, 291, 294, 384, 400, 402, 406, 411, 453, 454, 470, 482, 483, 492, 508, 968
6	Others	61	Agriculture, Husbandry, Food	43, 47, 56, 99, 111, 119, 131, 426, 449, 452, 460
		62	Amusement Devices	273, 446, 463, 472, 473
		63	Apparel & Textile	2, 12, 24, 26, 28, 36, 38, 57, 66, 68, 69, 79, 87, 112, 139, 223, 450
		64	Earth Working & Wells	37, 166, 171, 172, 175, 299, 405, 507
		65	Furniture, House Fixtures	4, 5, 30, 70, 132, 182, 211, 256, 297, 312
		66	Heating	110, 122, 126, 165, 237, 373, 431, 432
		67	Pipes & Joints	138, 277, 285, 403
		68	Receptacles	53, 206, 215, 217, 220, 224, 229, 232, 383
		69	Miscellaneous - Others	1, 14, 15, 27, 33, 40, 52, 54, 59, 62, 63, 84, 101, 108, 109, 116, 134, 135, 137, 150, 160, 168, 169, 177, 181, 186, 190, 199, 231, 236, 245, 248, 249, 269, 276, 278, 279, 281, 292, 300, 368, 404, 412, 428, 283, 289, 434, 441, 462, 503, 901, 903, 977, 984
7	Design	79	Design patents	Dxx
8	Plant	89	Plant patents	PLT

Source: Hall et al. (2001) plus own additions based on new technology codes