Colocation of Production and Innovation: Evidence from the United States*

Teresa C. Fort[†] Wolfgang Keller[‡] Peter K. Schott[§] Stephen Yeaple[¶]
Nikolas Zolas[‖]

December 15, 2020

Preliminary and Incomplete!
Comments Welcome!

Abstract

Manufacturers perform the majority of US patenting and R&D. The decades-long decline of US manufacturing employment raises concerns that US innovation will falter. We investigate the relationship between between physical production and innovation by constructing a new dataset linking all US firms and their establishments to location geocodes and innovative activities. Preliminary results indicate that while firms with both manufacturing and innovation establishments exhibit higher patenting when these facilities are more spatially proximate, manufacturing firms' overall contribution to US innovation declines steadily and substantially over time. Moreover, cohorts of firms permanently exiting manufacturing in the 1990s and 2000s continue to patent at their prior pace.

*We thank Nathan Goldschlag and Shawn Klimek for helpful comments on the paper, Erica Fuchs for a constructive discussion at the Atlanta Fed Trade Conference, seminar participants and CREI and Dartmouth for useful questions and suggestions, Emily Greenman for help with the disclosure process, and Alex Schott for research assistance. Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the US Census Bureau. The Census Bureau's Disclosure Review Board and Disclosure Avoidance Officers have reviewed this data product for unauthorized disclosure of confidential information and have approved the disclosure avoidance practices applied to this release. DRB Approval Numbers CBDRB-FY20-P1916-R8726 and CBDRB-FY20-P1916-R8756.

[†] Tuck School at Dartmouth & CEPR & NBER, email: teresa.fort@dartmouth.edu.

[‡] University of Colorado & CEPR & NBER.

[§] Yale School of Management & CEPR & NBER.

[¶] Penn State University & NBER.

[‖] Center for Economic Studies US Census Bureau.

1 Introduction

Generating ideas and bringing them to life as new products and processes is critical for growth. Recent research suggests US innovation efficiency is in decline, as more and more resources are used to create fewer and fewer ideas (Bloom et al., 2020). Given the large number of patents historically granted to manufacturers, some attribute this downturn to the erosion of US manufacturing employment (Autor et al., forthcoming). Indeed, the emergence of China as the world's factory floor has heightened concerns in high-wage countries like the United States that innovation will suffer as physical production shifts toward lower-wage developing countries. Writing in the Wall Street Journal, Kota and Mahoney (2019) assert that "once manufacturing departs from a country's shores, engineering and production know-how leave as well, and innovation ultimately follows. It's become increasingly clear that 'manufacture there' now also means 'innovate there'."

Theoretically, the relationship between innovation and the loss of manufacturing employment depends upon how firms respond to the shock causing the decline, and the nature of the complementarity between manufacturing and innovation. Increased import competition, for example, may lead some manufacturing firms to shrink or exit, so that production and innovation activities decrease in tandem. Alternatively innovation may rise if the resources formerly assigned to physical production are reallocated to the development of new products or production processes, both within and across firms. Such re-assignment may be more likely among "frontier" firms which possess greater capacity to either innovate away from foreign competitors (Aghion et al., 2005) or exploit offshoring opportunities (Boler et al., 2015; Bernard et al., 2020). Innovation efficiency may be further enhanced if there are gains to specialization in innovation (Arkolakis et al., 2018).

If complementarities between innovation and production workers are strong, however, declining manufacturing employment may reduce knowledge-worker efficiency, inducing an overall decrease in innovation (Naghavi and Ottaviano, 2009; Pisano and Shih, 2012).[1] Such complementarities may operate within the borders of a region, within the boundaries of a firm, or both. If face-to-face interactions between production and innovation workers affect innovators' output, then the geographic colocation of these activities may be critical, regardless of the firm boundary. Alternatively, if management facilitates and directs the flow of information across these two groups of workers, then the presence of both activities within the firm may increase innovative output, regardless of their geographic location. Finally, if face-to-face interactions matter more within firms, geographic colocation of the two activities within firms will be most valuable.

In this paper, we investigate the link between physical production and innovation using a novel dataset that matches the universe of US firms' and their US establishments to patent grants and R&D expenditures. In the first part of the paper we use these data to provide the first comprehensive portrait of US innovation spanning the decline of US manufacturing employment. Surprisingly, we find that while firms with manufacturing establishments, *i.e.*, manufacturing firms (MFs), historically account for the majority of innovation, their share has declined steadily and substantially over time. In 1977, MFs account for 91 percent of US patents and 99 percent of US R&D expenditures. By

[1] The effect of declining manufacturing employment on innovation may also be different if it is due to technical change and automation. Fort et al. (2018) discuss the roles of trade and technology in the evolution of US manufacturing.

2016, those shares had fallen by almost half, to 54 and 59 percent.

Decomposing firms outside manufacturing into two groups – never-manufacturing firms (NMFs) and former-manufacturing firms (FMFs) – we find that both provide considerable contributions to the growth of US innovation.[2] By 2016, NMFs and FMFs account for 28 and 18 percent of all US patent grants, up from a combined share of 9 percent in 1977. Moreover, patents per worker, an often-used measure of innovation efficiency (Griliches, 1994; Kortum, 1997), rise for both groups of firms over time. Indeed, we find that NMFs' patents per worker are roughly equal to those of MFs across the sample period, while those of FMFs are about half as high. FMFs' that exit manufacturing between 2007 to 2011 are particularly interesting since they not only exhibit dramatic growth in their patenting intensity after exiting, but also account for a concurrent surge in *FMF* imports from China. These trends suggest this group of firms may have exploited low-cost production opportunities in China for physical assembly while continuing to innovate in the United States.[3]

On the face of it, the growth of patenting and R&D expenditures by firms without manufacturing facilities presents a challenge to the idea that innovation depends on the presence of physical production, at least within the firm. We investigate this relationship more formally via a series of descriptive OLS panel regressions of patenting on a series of dummy variables capturing whether firms have manufacturing or innovation-related establishments – denoted M and P – or both.[4] We estimate these regressions with and without firm fixed effects, and find that firms with manufacturing plants have 1.5 to 3.7 log points more patent grants depending on the specification, while firms with an innovation plant have 0.5 to 1.7 log points more patents, both relative to firms with no such plants.

The more striking result is the estimated coefficient on the interaction of the M and P indicators, which suggests that firms with both types of plants have 66.5 log points more patents relative to firms without either of these establishments. Exploiting variation within firms over time, the interaction also suggests that firms have about 15 log points more patents in those periods during which they have both types of plants. Results are qualitatively similar for firms' manufacturing and processing patents, as well as for the citations received by their overall patent portfolio, an often used measure of patent quality (Trajtenberg, 1990). Together, these descriptive regressions suggest that manufacturing and innovation may be complementary within firm boundaries via one of the channels noted above.

To get a better sense of the nature of this complementarity, and the potential effects of the decline in US manufacturing employment on US innovation, we focus on firms with both M and P establishments – denoted MP firms – and investigate the relationship between their patenting and the distance between them. Toward that end, we identify the latitudes and longitudes of all US establishments using information contained in Census datasets. For each firm, we then measure the minimum distance between its M and P plant pairs. Using these new data, we find that MP firms'

[2] NMFs in a given year do not have a manufacturing plant in that or any previous year. FMFs in that year do not have a manufacturing establishment, but have encompassed such a plant in prior years.

[3] Greenland et al. (2020) offer a useful example along these lines from the electronics industry. Apple and Gateway both produced personal computers for the US market prior to the easing of US restrictions on Chinese imports in 2000. Apple had been making extensive use of Asian suppliers prior to this liberalization, and went on to benefit from further offshoring to China and focusing on product creation. Gateway, by contrast, focused on producing solely within the United States, and ceased operation a few years after the liberalization.

[4] As discussed in greater detail in Section 3.4, we refer to innovation establishments as P plants because they largely provide Professional Services, such as Scientific Research and Development Services (NAICS 5417).

whose M and P establishments are closer exhibit greater patenting. Coefficient estimates from our preferred specification, which includes firm fixed effects, indicate that when firms' closest M and P plants are within 5 miles of each other, their patent grants are 12 log points higher relative to when they are more than 60 miles apart, *i.e.*, more than an hour's drive away. Intuitively, we find that when firms' closest M and P plants are between 5 and 60 miles apart, the magnitude of this relationship is a bit smaller, at 8 log points. Here, too, results are qualitatively similar for citations as well as manufacturing and processing patents.

In a future draft, we plan to match the locations of all inventors granted patents to the nearest location in which the firm receiving the patent grant has a plant. We will then use a series of firm-region-year panel regressions to examine whether the patenting that takes place within MP firms occurs in regions where both M and P plants are present.[5] This analysis will speak to the mechanisms underlying the descriptive results summarized thus far, and in part address the endogeneity of firms' location decisions. More broadly, we plan to examine the intensive- and extensive-margin adjustments that give rise to changes in M-P plant distances over times, *e.g.*, whether they are driven by firm or plant entry and exit, or industry switching at the plant level, and how such changes relate to import competition and firms' offshoring decisions (e.g., as in Bernard et al., 2006; Bloom et al., 2019). We will also analyze the extent to which colocated establishments' patents differ from those of other plants, and whether across-firm M and P proximity relates to innovative output.

This paper makes three main contributions to the literature. First, we add to a large body of work that studies the economic geography of innovation and manufacturing. One strand of research studies how much and why manufacturing industries are agglomerated (Ellison and Glaeser, 1997; Duranton and Overman, 2005) and coagglomerated (Ellison et al., 2010). Related work documents strong patterns of spatial concentration of employment and establishments that focus on knowledge creation (Buzard et al., 2017; Davis and Dingel, 2019; Buzard et al., forthcoming), and show that this innovation clustering is distinct from production agglomeration (Audretsch and Feldman, 1996) motives. These papers emphasize the importance of localized knowledge spillovers (e.g., as in Jaffe et al., 1993) as an explanation for this agglomeration, and especially for unconventional innovations (Berkes and Gaetani, 2020). We build on these two strands of the literature by studying the extent to which physical production and innovation are spatially proximate. Several recent empirical papers analyze whether and how innovation is affected by its proximity to manufacturing (Tecu, 2013; Lan, 2019; Delgado, 2020). Ours is the first study to span the rise and fall of US manufacturing from 1977 to 2016, to analyze the spatial distribution of US manufacturing firms' innovation and manufacturing establishments using newly developed plant-level geocode information for this period, and to estimate how changes in the distances between these establishments relate to changes in a firm's patenting.

This paper also contributes to a literature that studies the evolution of manufacturing in high-income countries. A large body of work has documented a role of trade, and imports from China in particular, in the decline of US manufacturing (Autor et al., 2013; Pierce and Schott, 2016), and patenting (Autor et al., forthcoming). There is mounting evidence, however, that the firms that

[5] We will also use these data to assess whether manufacturing establishments themselves seem to patent, i.e., whether there is colocation of innovation and production within an establishment.

account for the majority of the decline in US manufacturing employment do not exit, and instead reorient towards non-manufacturing activities such as retail and business services (Fort et al., 2018; Ding et al., 2019; Bloom et al., 2019). This reorientation of manufacturers away from production towards innovation is consistent with evidence from other countries (Bloom et al., 2016; Bernard et al., 2017) and the emergence of factory-less goods producing firms that design goods and purchase contract manufacturing services, often from other countries (Bernard and Fort, 2015; Kamal, 2018).[6] We contribute to this work by showing that it is critical to include non-manufacturing firms in an analysis of US innovation, since there is substantial growth in patenting both by firms that decrease their manufacturing employment (sometimes all the way to zero), as well as by firms that never manufacture. Non-manufacturing firms' share of aggregate patents grows from 9 percent in 1977 to 46 percent in 2016.

Finally, we contribute to a growing literature on the relationship between offshoring and innovation. Theoretical work shows that offshoring may decrease innovation if there are strong complementarities between production and innovation (Naghavi and Ottaviano, 2009), or increase innovation if it lowers the opportunity cost to innovate (Rodríguez-Clare, 2010) and results in gains to specialization (Arkolakis et al., 2018). Fuchs and Kirchain (2010) find that offshoring to low-wage countries may also change the feasible set of goods that can be produced, which in turn will influence how firms direct their R&D efforts. There is less direct evidence on whether a firm's decision to relocate its physical production to a foreign location decreases its domestic innovation efforts and output. Bilir and Morales (2020) estimate that 20 percent of the returns to US multinational enterprises' R&D investments take place in their foreign affiliates, suggesting that innovation in the US leads to gains in foreign markets. Boler et al. (2015) document the presence of strong complementarities between R&D and imported imports that operate through a scale effect. Bernard et al. (2020) find that firms that exploit new production opportunities in low-wage countries increase their share and level domestic employment in research-related occupations and increase the quality of their domestically produced varieties. In contrast, Branstetter et al. (2020) document a relative decrease in patenting of specific products after a plausibly exogenous policy shock that allowed Taiwanese firms to offshore production of those goods to China. We build on this evidence by assessing whether and how geographic proximity within the United States relates to firms' patenting, since separating manufacturing and innovation across borders is less likely to be problematic if it is already geographically separated.[7]

The rest of the paper proceeds as follows. In Section 2 we describe how we construct the dataset used in Section 3 to describe US innovation since the late 1970s. Section 4 presents our analysis of patenting as a function of the spatial proximity of firms' M and P establishments. Section 5 will present our firm-region regressions in an upcoming draft. Section 6 concludes.

[6] This new type of innovation and production process is likely facilitated by the advent of information and communication technologies that facilitate production fragmentation, as in Fort (2017).

[7] In future drafts, we will also use firms' direct imports to assess how they relate to changes in its patenting, and to measure the extent to which traditional measures of import competition include direct imports by US manufacturing firms that patent. If these flows are large, then it is possible that the aggregate industry measures of Chinese import competition may actually capture US firms' offshoring decisions rather than exogenous increases in Chinese productivity that differ across sectors. For example, the large surge in Computer and Electronics imports from China may reflect offshoring by US firms that are specializing in innovation in response to low-wage production opportunities in China.

2 Data

In this section we describe how we construct a new dataset that tracks the employment, pay, production, location, and innovation of US firms and establishments.

2.1 US Firms and Establishments

We use the US Census Bureau's Longitudinal Business Database (LBD) initially assembled by Jarmin and Miranda (2002) to track the employment, pay, and industry of all private, non-farm US establishments from 1977 to 2016 annually. An establishment is a single physical location where business transactions take place and for which pay and employment are recorded.[8] The LBD contains a longitudinally consistent establishment identifier (*lbdnum*), as well as a firm identifier ($firmid$) that captures all of the establishments that are under common ownership or the control in a given year. We use $firmid$ to aggregate establishments to firms each year, and follow Ding et al. (2019) in implementing a series of corrections to this identifier over time.[9]

We track the unique 6-digit NAICS industry of each establishment in each year using the NAICS industry codes developed by Fort and Klimek (2018). These codes represent the major activity of the establishment. We use these codes to identify the mix of industries in which multiple-establishment firms operate, and to assign each firm principal six, three, and 2-digit NAICS codes according to its largest shares of employment and pay, respectively, at those levels of aggregation.[10]

A significant contribution of this paper is to assign a latitude and longitude (henceforth "geocodes") to all establishments in the LBD from 1977 to 2016. These codes allow us to document where manufacturing and innovation-related activities take place, and to analyze the extent to which they are spatially colocated. More broadly, they enable us to calculate the distance between any two establishments within a firm, as well as the distance between the establishments of different firms.

The Census Bureau's Business Register (BR) has geocodes for the majority of establishments starting in 2007. These geocodes are assigned to the Census block of the physical address of the establishment and are therefore the same for all establishments located in the same block. As these assignments were initially devised to cover residential zones, they are more apt to be missing for addresses in commercial areas. We use these geocodes for an establishment whenever they are available. In earlier years, and in cases where geocode information is missing from the BR, we attempt to recover a geocode by entering an establishment's street address and zip code into ArcGIS software. If that fails, we use the establishment's zip code from the LBD to assign the zip code's centroid latitude and longitude to the establishment.[11] This procedure yields a unique geocode for all establishments

[8]Technically, this information is reported at the firm (EIN) level and split to establishments by Census. This split is noisier in years between Economic Censuses.

[9]Census' $firmid$ can break spuriously over time for a number of reasons, such as when a firm transitions between having a single versus multiple establishments. We discuss the corrections suggested by Ding et al. (2019) in greater detail in Appendix Section B.

[10]In the cases where there is no variation in n-digit NAICS sectors within $(n-1)$-digit roots, we replace the n-digit codes with $(n+1)$-digit codes. For example, NAICS sector professional services (54) contains no variation at the 3-digit level – they are all 541. In that case, we use the $541x$ codes in place of 3-digit codes.

[11]Shifting zip code boundaries over our sample period present a challenge, as the centroid geocode may change with the zip code's borders, inducing spurious movement of establishments. We address this issue by assigning establishments

that is constant across all the years we observe the establishment. Additional details on our matching procedure are in Appendix Section A.

We augment the geocoded LBD with additional information from the Economic Censuses (ECs) of Manufacturing (CMF), Wholesale (CWH), Retail Trade (CRT), and Services (CSR), which are collected in years ending in "2" and "7", henceforth referred to as "Census" years. The EC data provide establishment-level measures of sales for all sectors. In addition, each Census has industry-specific information on additional variables, such input purchases (CMF, CWH), auxiliary status and industry served (CSR), number of products (CMF, CWH), and employment and wage bills by production versus non-production workers (CMF, CWH).[12] In 2007, the CMF and CWH also provide information related to establishments' primary activity, and whether or not establishments perform in-house design or purchase contract manufacturing services.

We also link the geocoded LBD to information on firm trade available from the Longitudinal Firm Trade Transactions Database (LFTTD). These data are based on US Customs transactions and capture all import and export transactions by 10-digit Harmonized System (HS10) product categories and source or destination country. The trade data are available from 1992 to 2018 and are collected by Employer Identification Number (EIN). The LFTTD maps EINs to *firmid*s.[13] As a result, we observe trade at the firm, but not the establishment, level.

2.2 Measures of Innovation

We measure US innovation using patents and research and development ($R\&D$) expenditures from 1977 to 2016.[14] We link these data to the geocoded LBD to construct a new dataset tracking the spatial distribution of firms' innovation activities.

US patent data are from the publicly available US PatentView (USPV) dataset produced by the US Patent and Trademark Office (USPTO).[15] USPV data provide the application date, grant date, assignee (i.e., firm) name, type and address (city, state and country), inventor name(s) and address(es) (city, state or country), forward and backward citations, and patent technology class for every patent *granted* in the United States between 1976 and 2020. While the vast majority of patents contain information allowing us to link them to firms, such as their name, city and state, approximately 14 percent of patents are missing these data. Our analysis focuses on patent grants because these are patents deemed to be credible contributions to knowledge by patent examiners, and because of the greater availability of identifier information for patent grants versus applications.[16]

We combine the USPV data with our geocoded LBD using the assignee name and city-state address information of the assignee and inventor(s), which we match to firm name and address in

the mean geocode across our sample period.

[12] Auxiliary establishments provide support functions for other establishments within a firm.

[13] Canadian exports are only collected with firm name information. For details on the original construction of the LFTTD, and as well as a more recent update, see Bernard et al. (2007) and Kamal and Ouyang (2020).

[14] We plan to expand the set of innovative activities we examine to other items, such as trademarks, in a future draft.

[15] These data can be downloaded from the USPTO's website at `http://www.patentsview.org`. The patent information used in this paper was accessed on May 19, 2020.

[16] We are investigating whether recent data assembled by the USPTO permit greater matching of patent applications to US firms.

the Census Bureau's Business Register (BR). We note that we match patent grants to firms in their *application* year, as the interval between patent application and grant can stretch several years and our interest is in the firm characteristics that give rise to innovation. Overall, we match 64 percent of the nearly 3 million patents granted to US firms between 1977 and 2016 to a specific firm and application year in the LBD. These match rates are relatively constant over time, varying between 60 and 67 percent across years. Appendix Section C provides additional information on the patent data and our matching algorithm.

As just noted, an important feature of granted patents is the time lag between patents' application and grant dates. Comparing the left panel of Appendix Figure A.1 to Appendix Figure A.3, it is clear that the count of granted applications by application year declines at the end of the sample due to this lag, as opposed to a decline in patenting. The first application year appearing in our matched data is 1972, linked to firms in the LBD in 1976. Our analysis begins in 1977, but we use patent grants applied for before this period to estimate patent stocks.

The USPTO classifies patents into specific technology groups according to cooperative patent classification codes (CPC), jointly developed by the USPTO and the European Patent Office (EPO).[17] We are able to link the CPC technology classes to 6-digit NAICS industries using the Algorithmic Links with Probabilities (ALP) procedure developed by Lybbert and Zolas (2014) and updated in Goldschlag et al. (2019). This concordance links particular technologies to industries probabilistically based on the underlying keywords that describe the domain of goods and services they represent.[18] For instance, patents related to Magnetic and Optical Media manufacturing (NAICS 3346) link to CPC classes Information Storage (CPC G11) and Computing, Calculating and Counting (CPC G06). Likewise, software products in Computer Systems Design and Related Services (NAICS 5415), particularly Custom Computer Programming Services (NAICS 541511), link to technology class Data Processing Systems or Methods (G06Q) and other technology classes within Computing, Calculating and Counting (CPC G06). We identify processing patents using data from Ganglmair et al. (2020), who exploit the standardized language of patent claims to identify the grammatical structure and keywords associated with process versus product patents.

In addition to matching the USPV data to firms in the geocoded LBD, we exploit the inventor addresses to match each patent to a firm-city-state, and thereby county Federal Information Processing Standards ($FIPS$) codes and Commuting Zones (CZs). For single-establishment firms, this is identical to matching the patent to a firm. However, the majority of patents are granted to large, multi-establishment firms that are active in multiple locations. For these firms, we assess whether the patent inventor is in the same city-state as one (or more) of the firm's establishments. If it is, then we assign the patent to that city-state. If the firm has no establishments in the inventor's city-state, we then assign the patent to the closest city-state in which the firm has one (or more) establishments.

[17]The nine major CPC categories are: Human Necessities; Performing Operations and Transporting; Chemistry and Metallurgy; Textiles and Paper; Fixed Constructions; Mechanical Engineering, Lighting, Heating, Weapons and Blasting; Physics; Electricity; and Other. Appendix Figure A.2 provides an annual breakdown of patent grants by these major groupings.

[18]More precisely, the algorithm combs through the abstracts and titles of a global set of patents for keywords specifically developed to describe each NAICS classification, before applying a filter and re-weighting the matches to minimize Type I and Type II errors.

Further details on the matching procedure are in Appendix Section C.3. Exploiting the inventor addresses thus allows us to identify which of the firms' establishments are most closely associated with patenting activity.[19]

We measure US firms' research and development (R&D) expenditures using two Census Bureau surveys: the Survey of Industrial Research and Development (SIRD) and the Business R&D and Innovation Survey (BRDIS), which cover years 1977 to 2007, and 2008 to 2016, respectively. The SIRD and the BRDIS record surveyed firms' overall R&D spending as well as various breakdowns of this total, *e.g.*, by foreign versus domestic expenditure, as well as basic versus applied and development. The BRDIS and SIRD also record the number of scientists, engineers, and technical workers employed by the firm. It is important to note that we observe R&D expenditures only for surveyed firms. Traditionally, the SIRD focused on large manufacturing firms, though the scope for the two surveys has increased over time so that both surveys sample a nationally representative set of firms.[20] In terms of the spatial distribution of R&D, the surveys break down R&D expenditures by state, thus providing some geographic variation for large, multi-unit firms that span different US states.

3 An Overview of US Innovation

In this section we provide an overview of US firms' innovative activities over the past 45 years, a period which spans the decline of US manufacturing employment from its peak of 19.1 million workers in 1979 to 11.5 million in 2016.

The left panel of Figure 1 reports the number of patents subsequently granted to firms located in the United States by their application year. The right panel plots firms' total worldwide R&D expenditures. Patent grants are flat in the early years of the sample, and then again between 2000 and 2007. As noted in the last section, the decline in granted patents starting in 2014 is an artifact of the often multi-year lag between patents' application and grant dates, which averages 3 years but can last up to 7 years.

3.1 Innovation by Manufacturers and Non-Manufacturers

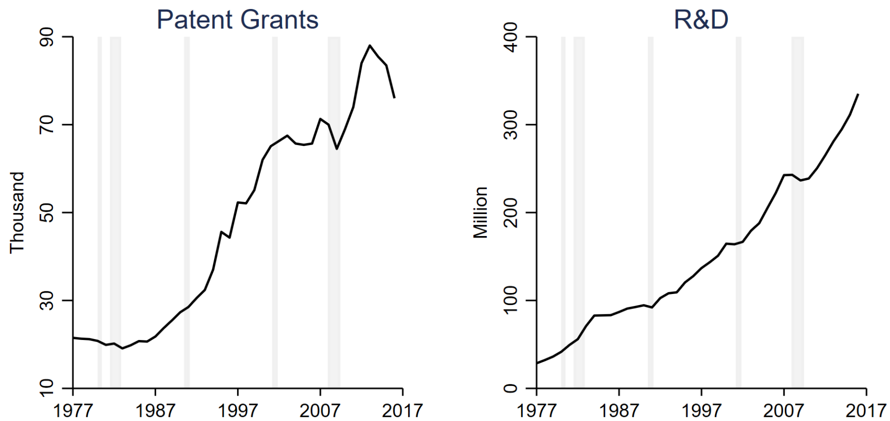
To analyze the evolution of US innovators, and the role of manufacturers in particular, we categorize firms in year t into three mutually exclusive groups: manufacturing firms (MFs), former manufacturing firms (FMFs), and never-manufacturing firms (NMFs). A MF in year t is defined as a firm that includes a manufacturing establishment in that year. A FMF in year t, by contrast, is a firm without any manufacturing plants in that year, but which did encompass at least one manufacturing plant in some prior year. Lastly, a NMF in year t is a firm which has not had a manufacturing establishment up to and including year t.[21] Note that these definitions are with respect to the sample

[19] Figure A.6 of the Appendix uses publicly available data to provide an example of the distribution of inventors of Bristol Myers Squibb Patent 10167343 around that firm's NJ facilities.

[20] Sampling is limited to firms with 10 or more employees. Firms with two consecutive years of zero R&D expenditures are permanently dropped from the survey, while firms with at least one R&D establishment (NAICS 5417) are surveyed with certainty.

[21] NMFs may include a manufacturing plant in some future year. In practice, the number of such firms is very small compared to the overall number of NMFs in each year.

Figure 1: US Innovation

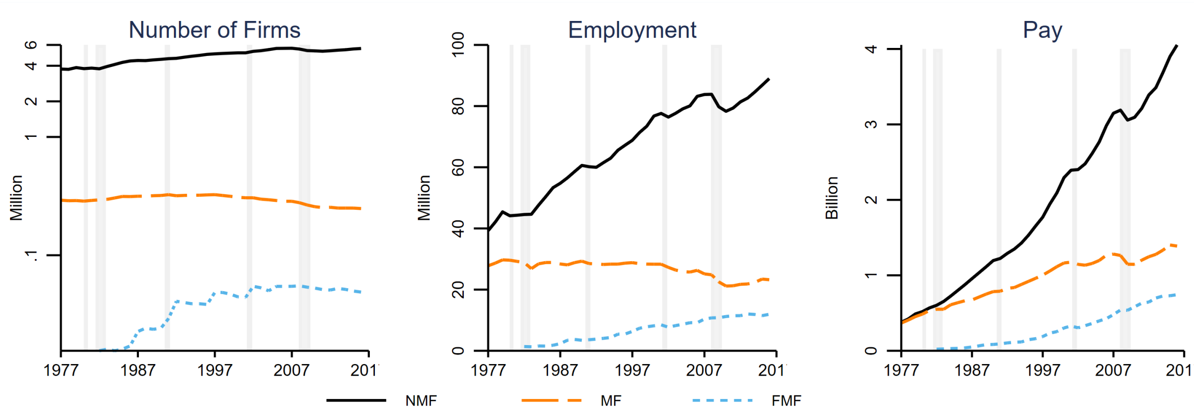


Source: LBD, BR, USPV, SIRD, BRDIS, and author's calculations. Left panel displays number of patents subsequently granted to firms by their application year. Right panel displays total R&D expenditure of US firms. In both cases, totals are for the set of firms for which a match between the patent and R&D data could be found (see main text for further detail). Vertical bars represent NBER-dated recessions.

period, as we have no information about firms prior to 1977. As a result, the number of FMFs in 1977 is zero by definition. Due to the very small number of FMFs in the initial years of our sample period, we do not break out FMFs from NMFs until 1982 in the results to follow.

The three panels of Figure 2 provide a breakdown of US firms by type in each year, as well as their total employment and payroll. As indicated in the figure, NMFs represent the vast majority of firms, but far smaller shares of total employment and payroll. Manufacturing firms' total employment (including workers at their manufacturing and non-manufacturing establishments) is relatively flat until the 2000s, at which point it declines. FMF employment, by contrast grows steadily, even as the the number of FMF firms' plateaus at the end of the period.

Figure 2: US Firms by Type

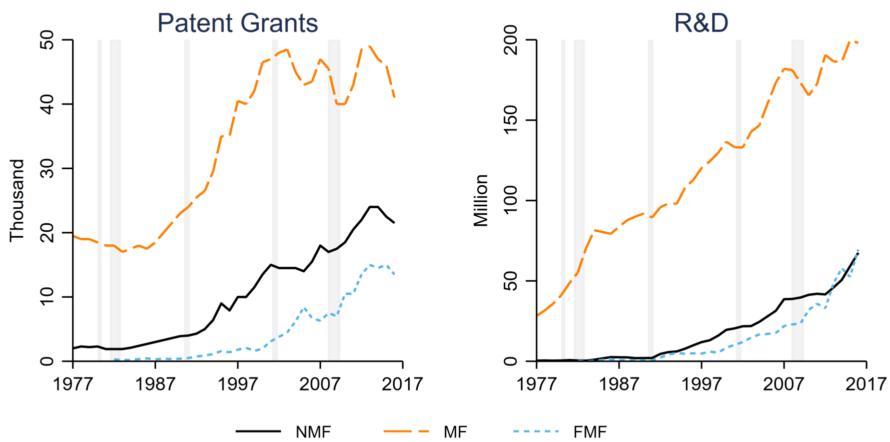


Source: LBD and author's calculations. Left panel displays numbers of US firms by type using a log scale. Middle and right panels display their employment and nominal payroll. Firm types are non-manufacturing firms (NMFs), manufacturing firms (MFs) and former manufacturing firms (FMFs), as defined in the main text. Data for FMFs are suppressed prior to 1982.

Figure 3 provides a breakdown of US patenting and R&D expenditures by type of firm. As indicated in the left panel, MFs account for the majority of granted patents and R&D expenditures in all years of the sample, though their share of these activities declines substantially over time, from 91 percent in 1977 to 54 percent in 2016. The decline in R&D expenditures, in the right panel, is similar, from 99 in 1977 to 59 percent in 2016.[22]

A stark difference between the two panels is the overall flatness of granted patent growth among MFs between 2000 and 2010, a period book-ended by the 2001 and 2007 recessions. Patent growth by NMFs and FMFs, by contrast, continues apace during this period. NMFs account for 9.0 percent of US patents in 1977 and 28 percent in 2016 – a rise of 211 percent. FMFs account for zero patents in 1977 (by definition) and 18 percent of US patents in 2016, suggesting that at least for some firms, the continued presence of manufacturing activity within the firm (and within the US) is not essential for successful innovation.

Figure 3: Innovation by Firm Type



Source: LBD, BR, PatentsView, SIRD, BRDIS and author's calculations. Left and right panels report patents granted to US firms by their application year and total research and development (R&D) expenditures, by type of firm. Firm types are non-manufacturing firms (NMF), manufacturing firms (MF) and former manufacturing firms (FMF), as defined in text. Data for FMF are suppressed prior to 1982.

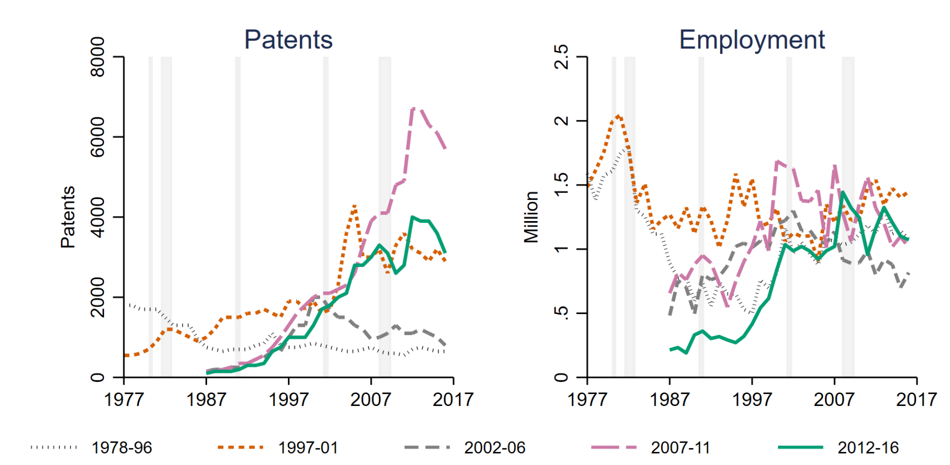
We provide further context for FMFs in Figure 4, which presents a breakdown of these firms' patent grants according to the year in which they permanently exit manufacturing. In the figure, FMFs are assigned to one of five cohorts depending upon the year in which they drop their last manufacturing establishment, e.g., between 1977 and 1996.[23] Though they keep patenting, the number of patents granted to firms that permanently exit manufacturing prior to 1997 declines over time. In contrast, three of the four cohorts of firms that exit manufacturing after 1997 exhibit relatively strong patent growth, both prior to and after shedding their manufacturing plants. The exception is the 2002 to 2006 cohort, for whom patenting declines after 2001. Patent growth among the 2007 to

[22] In unreported results, we find that manufacturing firms account for a similarly declining share of manufacturing patents. Appendix Section C provides a decomposition of US patents by NAICS sector. Manufacturing patents' share of overall patents falls from 92 in 1977 to 84 percent in 2016. Among publicly US traded firms in 2007, Autor et al. (forthcoming) find that manufacturers account for more than two-thirds of US corporate patents and R&D spending.

[23] Patent counts prior to 1987 are suppressed for FMFs in the final three cohorts. Note that these cohorts are not "balanced" in the sense that they contain a mix of firms that survive versus subsequently exit.

2011 cohort is particularly strong. The right panel of Figure 4 reveals that the overall employment of each cohort is broadly similar, indicating that variation in their patenting is not simply a reflection of differential growth.

Figure 4: Patenting and Employment by Permanent FMF Cohort



Source: LBD, LFTTD, PatentsView, and author's calculations. Panels report patents granted to permanent former manufacturing firms by cohort. Permanent FMFs are assigned to one of five cohorts depending upon the year in which they dropped their last manufacturing establishment during the sample period. For example, the lines for the 1977 and 1996 cohorts represent the patents granted to and the employment of all firms for which the last year in which they are observed to have a manufacturing establishment is 1977 to 1996. Patent counts prior to 1987 are suppressed for FMFs in the final three cohorts. Cohort lines include firms that exit prior to the end of the sample period.

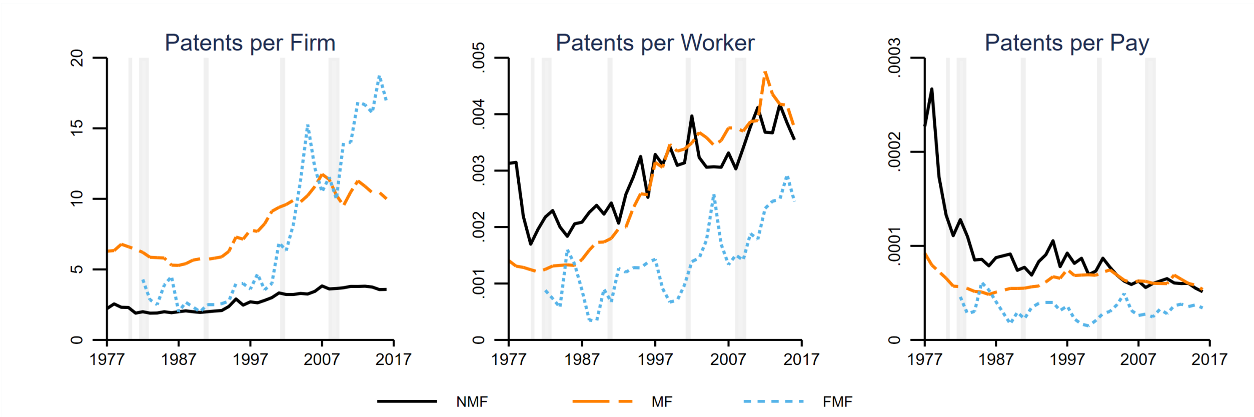
Patents per worker is an often-used indicator of innovation efficiency (Griliches, 1994; Kortum, 1997). For firms with patent grants, Figure 5 displays the number of grants per firm, per total employment, and per total nominal payroll. We find in the left panel that while MFs historically have had the largest number of patents per firm, they are overtaken by FMFs in the early 2000s. Comparison of the panel with Figure 4 reveals that this growth is driven by the cohort of FMF firms that leave manufacturing beginning in 1997.

The middle panel of Figure 5 shows that patents per worker are similar and rising over time for both MFs and NMFs. Patent efficiency for FMFs, while lower, also rises over time. These trends contrast with recent research suggesting ideas are getting harder to find. Bloom et al. (2020), for example, provide a series of examples in which innovation per worker is declining.

In the final panel of Figure 5, we exploit our ability to observe firms' payroll to compute patent grants per dollar of pay for each type of firm. As indicated in the figure, trends are either flat or decreasing after initially stark declines in the late 1970s and early 1980s.

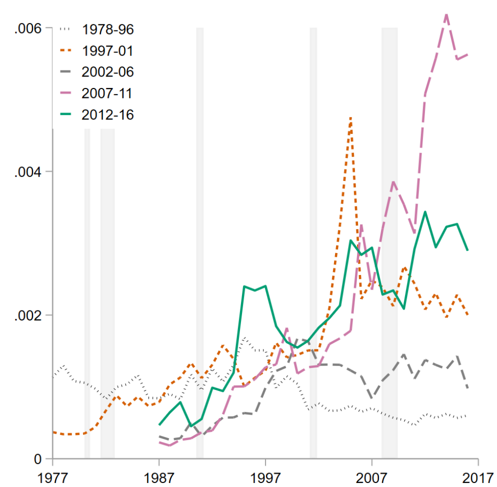
Finally, Figure 6 plots the patent efficiency of firms permanently exiting from manufacturing by their exit cohorts. Here, too, the later cohorts stand out in terms of seeing their efficiency continuing to rise after exit.

Figure 5: Patent Efficiency



Source: LBD, LFTTD, PatentsView, and author's calculations. Panels report patents per firm, per total employment, and per nominal payroll, by firm type. For each year, sample is restricted to firms with patent grants. Firm types are non-manufacturing firms (NMF), manufacturing firms (MF) and former manufacturing firms (FMF), as defined in text.

Figure 6: Patent Efficiency by Permanent FMF Cohort



Source: LBD, LFTTD, PatentsView, and author's calculations. Panels report patents granted per total employment to permanent former manufacturing firms by cohort. Permanent FMFs are assigned to one of five cohorts depending upon the year in which they dropped their last manufacturing establishment during the sample period. For example, the lines for the 1977 and 1996 cohorts represent the patents granted to and the employment of all firms for which the last year in which they are observed to have a manufacturing establishment is 1977 to 1996. Patent counts prior to 1987 are suppressed for FMFs in the final three cohorts. Cohort lines include firms that exit prior to the end of the sample period.

3.2 Trade and Innovation

The decline of US manufacturing employment has been accompanied by an increase in US imports, particularly from China (Autor et al., 2013; Pierce and Schott, 2016). Autor et al. (forthcoming) link this increase to a decline in US innovation, though Bloom et al. (2016) find the opposite relationship between Chinese import penetration and innovation in a study of European manufacturers. In this section, we describe how US trade varies across firm's type and patenting status.

The eight panels of Figure 7 focus on overall US imports and exports, as well as US imports from and exports to China. Each row of panels focuses on a different trade flow. Within each row, panels on the left examine trends among non-patenting firms while those on the right are for patenters. Panels in the same row have the same scale for their y-axis. All panels express trade in terms of billions of US dollars.

Three trends stand out across the panels. First, US trade predominantly flows through MFs that patent. As evident in the top four panels of Figure 7, aggregate imports and exports of non-patenting MF and NMFs are similar, while MF patenters' imports are about twice as large in 2016, and their exports are 3 times as large compared to non-patenters. The bottom four panels display similar patterns for Chinese imports and exports, with the notable exception of NMFs' Chinese imports, which are similar in magnitude to MFs' Chinese imports.

Second, in stark opposition to the trends for MFs, NMFs that do not patent trade *more* than NMFs that do patent. In fact, the high share of NMF imports from China is driven by non-patenting firms. This result may not be surprising since wholesalers and retailers are particularly important for US imports from China (Bernard et al., 2010). MFs' Chinese imports, by contrast, are about twice as large for patenters, suggesting a potential complementarity between offshoring and innovation. The third row of Figure 7 highlights the fact that the majority of US imports from China are mediated by MFs and FMFs, and thus raises the possibility that aggregate measures of Chinese import competition may reflect offshoring by US manufacturers. We plan to investigate this possibility in a future draft.

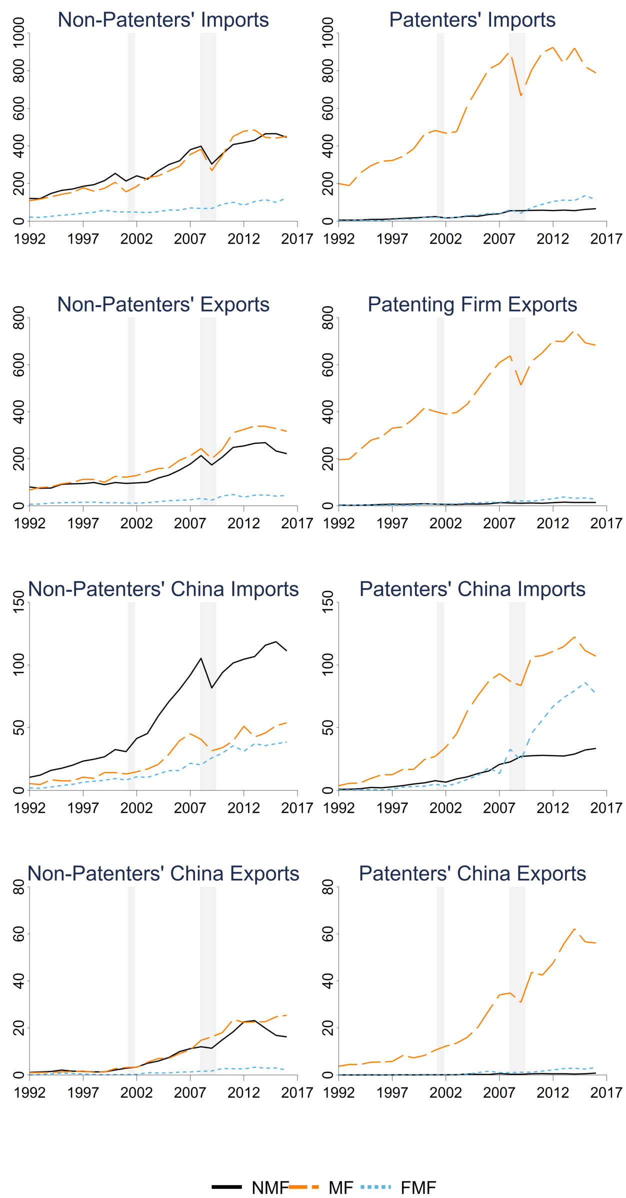
Finally, the third panel of Figure 7 shows that US imports from China by FMFs that patent surge starting in the mid-2000s. We examine these imports more closely in Figure 8, which plots overall US imports as well as US imports from China for the FMF cohorts described above. We find that FMFs' import growth is driven by the cohort that leaves manufacturing between 2007 and 2011. This cohort's total imports exhibit strong growth throughout the 1990s, and their Chinese imports accelerate after 2002. This cohort also exhibits the strongest growth in patent grants in Figure 4. Together, these trends suggest firms leaving manufacturing after the Great Recession may have begun relying on Chinese production for physical assembly while continuing to innovate. We return to this potential relationship below.

3.3 Industry Variation in Patenting

In a future draft, this section will document the distribution of patenting versus non-patenting firms' employment across 2-digit NAICS sectors. That analysis will demonstrate that several sectors are disproportionately large among patenting firms. Examination of these sectors' descriptions also provides further rationale for the view that they support innovation.

Based on the patenting firms' disproportionate employment shares and the NAICS industry descriptions, we classify establishments as innovation plants if they are in one of the following Professional and Technical Services (54) or Information (51) sectors: Architectural, Engineering, and

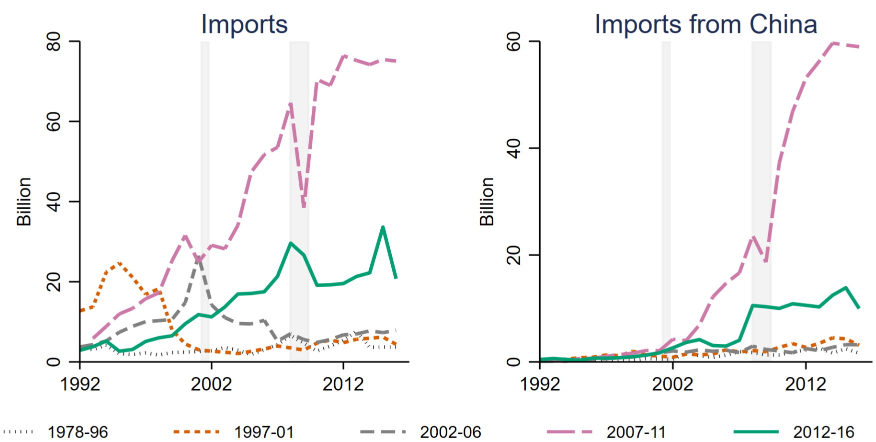
Figure 7: US Trade by Firm Type and Patent Status



Source: LBD, LFTTD, and authors' calculations. Left panels report aggregate US imports and exports as well as China imports and exports for firms that do not patent, by firm type. Right panels report analogous information for patenting firms. Firm types are non-manufacturing firms (NMF), manufacturing firms (MF) and former manufacturing firms (FMF), as defined in text. Data for FMF are suppressed prior to 1982.

Related Services (5413); Specialized Design Services (5414); Computer Systems Design and Related Services (5415); Management, Scientific, and Technical Consulting Services (5416); Scientific Research and Development Services (5417); Software Publishers (5112); Telecommunications (517); and Data Processing, Hosting, and Related Services (518). We also classify establishments in Corporate,

Figure 8: FMF Patenting Firms' Imports by Cohort



Source: LBD, LFTTD, and authors' calculations. Top panel reports aggregate US imports and exports by firm type. Bottom panel reports US imports from and exports to China by firm type. Firm types are non-manufacturing firms (NMF), manufacturing firms (MF) and former manufacturing firms (FMF), as defined in text. Data suppressed prior to the 2007-11 cohort in 1992.

Subsidiary, and Regional Managing Offices (551114) as innovation plants. Establishments that perform more than one core support function (e.g., Accounting and R&D) are classified as 551114, and these establishments often explicitly include R&D employment. Given the importance of Professional Services in this list of sectors, we refer to these innovation plants as "P" plants.

3.4 Patenting by MFs versus NMFs and FMFs

Our analysis thus far has shown that while manufacturing firms are important patenters, non-manufacturing firms increasingly contribute to aggregate US patent growth. We explore the relationship between patenting and firms' mix of manufacturing (M) and innovation (P) plants more formally via the following OLS regression:

$$\begin{aligned} ln(\widetilde{y}_{ft}) = \quad & \gamma_1 M_{ft} + \gamma_2 P_{ft} + \gamma_3 M_{ft} \times P_{ft} + \\ & \gamma_4 FMF_{ft} + \gamma_5 FMF_{ft} \times P_{ft} + \\ & \beta X_{ft} + \alpha_t + \alpha_r + \varepsilon_{frt}, \end{aligned} \tag{1}$$

where $\widetilde{y}_{it}$ represents an innovation outcome such as a patent count for firm f in year t, M_{ft} and P_{ft} are dummy variables indicating whether firm f has M or P establishments in year t, and $M_{ft} \times P_{ft}$ is the interaction between those dummies. In some of the specifications, we also include indicators to identify if a non-manufacturing firm is a former manufacture (FMF), and the interaction of $FMF_{ft} \times P_{ft}$, which identifies whether a FMF has P plants in that year. X_{ft} denotes a vector of time-varying firm characteristics such as employment size and age bins.[24] Finally, Equation (1)

[24]Following Haltiwanger et al. (2013), we measure firm age in year t as the age of the age of the oldest establishment in the firm in t. As data on plants are not available prior to 1977, plants present in that year are assumed to be born in that year.

includes year fixed effects to remove any macroeconomic drivers of patenting common across firms, and FIPS fixed effects for the firm's main county based on employment. We also estimate a variant of Equation (1) with firm fixed effects to assess whether a particular firm's patenting output varies with changes in its M and P status. We restrict the sample period to Census years between 1977 and 2012, which captures firm patenting activity through 2016.[25]

We estimate Equation (1) using a range of innovation outcomes for y_{it}, including the firm's overall number of granted patents, as well as its manufacturing and processing patent counts.[26] As our sample period is restricted to Census years, y_{ft} is computed as the sum of patents granted to firm f as of May 2019 (the date of our USPV download) for which the firm applied in years t through $t+4$. To account for potential variation in patent quality, we also examine patent citations. In that case, $\widetilde{y}_{ft}$ is the the total number of citations as of May 2019 earned by the patents granted to firm f that were applied for in years t through $t+4$.

As patenting is an unusual event and there are therefore many zeros in the data, we transform the patent and citation counts just described using an inverse hyperbolic sine function,

$$ln(\widetilde{y}_{ft}) = \ln\left(y_{ft} + (y_{ft}^2 + 1)^{1/2}\right). \tag{2}$$

Given the close correspondence between the natural log and this transformation, we will use "log" as a shorthand when referring to it, and note that the regression coefficients in Equation (1) may loosely be thought of as the percent difference in patent grants among firms with the noted attributes relative to the left-out group, *i.e.*, firms without either M or P plants.

We report the results from estimating Equation (1) via Ordinary Least Squares (OLS) in Table 1. Column 1 presents the estimates without firm fixed effects and shows that firms with a manufacturing plant have 3.7 log points more patents, while firms with an innovation plant have 1.7 log points more patents, both relative to firms with no manufacturing or innovation plants. The most striking result in Table 1 is the estimated coefficient on the interaction between M_{ft} and P_{ft}, which indicates that firms with both manufacturing and innovation in a given year have 67 log points more patents.

Columns 2 and 3 in Table 1 add the FMF indicator and the interaction between FMF and P. The estimates for these variables indicate that former manufacturing firms do patent relatively more than firms without M or P plants, but that this result is driven by FMF firms that have P plants. In fact, Column (3) shows that FMFs without P plants have 1.6 log points *fewer* patents, while FMFs with P plants have 23 log points more patents than firms without M or P establishments. Perhaps most surprising is the fact that the estimated coefficient on P in Column 3 is negative, large, and statistically significant, suggesting that firms with P establishments patent more only when they also have manufacturing, or had manufacturing in the past.[27]

Columns 4 to 6 in Table 1 present the results from estimating Equation (1) with firm fixed effects. Although the coefficient magnitudes are smaller, the same basic patterns are evident for firms with M

[25]As noted in Section 2, Census years end in 2 and 7.

[26]Ganglmair et al. (2020) exploit the standardized language of patent claims to identify the grammatical structure and keywords associated with process versus product patents. We thank them for generously providing their data.

[27]The full effect of being an FMF with a P plant in Column 3 is actually quite small, at 0.2 log points (*i.e.*, -0.213-0.016+0.231).

and P establishments. Firms tend to patent more when they have one or the other, and their granted patents are 15 log points higher in the years in which they have both. In contrast to the specifications without firm fixed effects, firms patent more when they are FMFs, relative to when they have no M or P plants. As in the cross-firm specifications, the relationship between FMF status and patenting is stronger in years in which the FMFs also have at least one P plant.[28]

Table 1: Patenting by Manufacturing versus Non-Manufacturing Firms

Dependent variable is $ln(Patents_{f,t:t+4})$: firm f's total patent grants from t to $t+4$

	(1)	(2)	(3)	(4)	(5)	(6)
M_{ft}	0.0374***	0.0376***	0.0365***	0.0149***	0.0179***	0.0174***
	(0.0003)	(0.0004)	(0.0003)	(0.0009)	(0.001)	(0.001)
P_{ft}	0.0172***	0.0172***	-0.213***	0.0047***	0.0047***	-0.0199
	(0.0004)	(0.0004)	(0.0206)	(0.0006)	(0.0006)	(0.0126)
MP_{ft}	0.665***	0.665***	0.707***	0.147***	0.147***	0.154***
	(0.0132)	(0.0133)	(0.0149)	(0.0082)	(0.0082)	(0.0090)
FMF_{ft}		0.0046*	-0.016***		0.0081***	0.0061***
		(0.0024)	(0.0015)		(0.0019)	(0.0016)
$FMF_{ft} \times P_{ft}$			0.231***			0.0247*
			(0.0206)			(0.0127)
$Workers_{ft}$						
10 - 99	0.0059***	0.0059***	0.0059***	0.0021***	0.0021***	0.0021***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
100 - 499	0.0518***	0.0518***	0.0521***	0.0226***	0.0225***	0.0226***
	(0.0009)	(0.0009)	(0.0009)	(0.0007)	(0.0007)	(0.0007)
500 - 4999	0.255***	0.256***	0.257***	0.112***	0.111***	0.111***
	(0.0050)	(0.0050)	(0.0050)	(0.0035)	(0.0035)	(0.0035)
5000+	1.505***	1.506***	1.515***	0.563***	0.562***	0.562***
	(0.0398)	(0.0398)	(0.0397)	(0.0216)	(0.0216)	(0.0216)
Age_{ft}						
5 - 9	-0.0017***	-0.0017***	-0.0017***	-0.0012***	-0.0012***	-0.0012***
	(0.0000)	(0.0000)	(0.0000)	(0.0001)	(0.0001)	(0.0001)
10+	-0.0030***	-0.0030***	-0.0030***	-0.0018***	-0.0018***	-0.0018***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
FIPS Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Firm Fixed Effects	No	No	No	Yes	Yes	Yes
R-squared	0.152	0.152	0.153	0.742	0.742	0.742
N (millions)	27	27	27	27	27	27

Source: LBD, BR, USPTO and authors' calculations. Table reports the results of estimating Equation 1 using the universe of firms in the US in Economic Census years 1977 to 2012 that are present in at least two Census years. This restriction ensures that the samples are identical across all specifications, including those with firm fixed effects, and does not alter the results. Dependent variable is the inverse hyperbolic sine transformation of the total patents granted to firm f between Census year t and $t+4$. Its mean and standard deviation are 0.0074 and 0.1360, respectively. M_{ft} and P_{ft} are dummy variables indicating whether firm f has manufacturing or innovation (P) establishments in Census year t. FMF_{ft} indicates whether firm f is a former manufacturing firm in year t. FIPS fixed effects capture the main FIPS code for a firm, based on its employment. Standard errors clustered by firm. Number of observations is rounded per Census disclosure guidelines.

Table 2 presents results from estimating Equation (1) for the firm's patent citations, manufacturing patents, and its processing patents. For each of these outcome variables, we present the results from including all covariates in Equation (1). Columns 1, 3, and 5 present estimates without firm

[28]In Column 6, the full effect of being an FMF with a P plant is 1 log point (*i.e.*, -0.02+0.006+0.024).

fixed effects, while columns 2, 4, and 6 include them. The primary and strongest message from the specifications with firm fixed effects is that firms with M and P plants have considerably higher patent citations, manufacturing patents, and processing patents than firms without M or P. In addition, firms tend to patent more in those years in which they have both of these types of establishments. Table 2 also shows that FMF firms with P plants tend to have more manufacturing and processing patents than firms without M or P, though fewer patent citations.

Table 2: Patent Citations and Counts by Manufacturing versus Non-Manufacturing Firms

Dependent variable is:

	$ln(Citations_{f,t:t+4})$		$ln(ManufPatents_{f,t:t+4})$		$ln(ProcessingPatents_{f,t:t+4})$	
	(1)	(2)	(3)	(4)	(5)	(6)
M_{ft}	0.0817***	0.0250***	0.0311***	0.0142***	0.0098***	0.0064***
	(0.0007)	(0.0024)	(0.0003)	(0.0011)	(0.0001)	(0.0007)
P_{ft}	-0.174***	0.188***	-0.213***	-0.0141	-0.135***	-0.0435***
	(0.0348)	(0.0231)	(0.0196)	(0.0115)	(0.0150)	(0.0094)
MP_{ft}	1.195***	0.167***	0.655***	0.139***	0.367***	0.0868***
	(0.0234)	(0.0163)	(0.0143)	(0.0083)	(0.0106)	(0.0063)
FMF_{ft}	-0.0391***	-0.0196***	-0.0116***	0.0033**	-0.0061***	0.0041***
	(0.0032)	(0.0032)	(0.0013)	(0.0014)	(0.0007)	(0.0010)
$FMF_{ft} \times P_{ft}$	0.211***	-0.184***	0.229***	0.0182	0.145***	0.0466***
	(0.0348)	(0.0232)	(0.0196)	(0.0116)	(0.0150)	(0.0094)
$Workers_{ft}$						
10 - 99	0.0129***	0.0027***	0.0047***	0.0016***	0.0019***	0.0006***
	(0.0002)	(0.0002)	(0.0001)	(0.0001)	(0.0000)	(0.0000)
100 - 499	0.113***	0.0330***	0.0422***	0.0181***	0.0148***	0.0074***
	(0.0017)	(0.0015)	(0.0008)	(0.0007)	(0.0005)	(0.0004)
500 - 4999	0.477***	0.156***	0.224***	0.0962***	0.115***	0.0520***
	(0.0086)	(0.0066)	(0.0047)	(0.0032)	(0.0031)	(0.0022)
5000+	2.411***	0.731***	1.390***	0.497***	1.013***	0.366***
	(0.0561)	(0.0353)	(0.0389)	(0.0201)	(0.0328)	(0.0173)
Age_{ft}						
5 - 9	-0.0035***	-0.0010***	-0.0014***	-0.0009***	-0.0008***	-0.0007***
	(0.0001)	(0.0001)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
10+	-0.0064***	0.0004**	-0.0025***	-0.0014***	-0.0014***	-0.0013***
	(0.0001)	(0.0002)	(0.0001)	(0.0001)	(0.0000)	(0.0001)
FIPS Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Firm Fixed Effects	No	Yes	No	Yes	No	Yes
R-squared	0.119	0.661	0.154	0.749	0.141	0.743
N (million)	27	27	27	27	27	27
Dep Var Mean & SD	0.0153, 0.2690		0.0063, 0.123		0.0029, 0.0797	

Source: LBD, BR, USPTO and authors' calculations. Table reports the results of estimating Equation 1 using the universe of firms in the US in Economic Census years 1977 to 2012 that are present in at least two Census years. This restriction ensures that the samples are identical across all specifications, including those with firm fixed effects, and does not alter the results. Dependent variables are the inverse hyperbolic sine transformation of total citations, manufacturing patents, and processing patents for patents granted to firm f between Census year t and $t+4$. M_{ft} and P_{ft} are dummy variables indicating whether firm f has manufacturing or innovation (P) establishments in Census year t. FMF_{ft} indicates whether firm f is a former manufacturing firm in year t. FIPS fixed effects capture the main FIPS code for a firm, based on its employment. Standard errors clustered by firm. We control for firm employment size and age categories. Observations rounded per Census disclosure guidelines.

Overall, the descriptive regressions presented in this section reveal that firms with M and P establishments patent considerably more than other types of firms. While FMF firms also have higher patenting output when they have a P establishment, this interaction is considerably smaller than the relationship between patenting and having both M and P plants. Together, these trends suggest a potential complmentarity between innovation workers and manufacturing activity within the firm, one that may persist even after the firm exits physical production. We investigate this potential complementarity further in the next section by taking into account the spatial proximity of firms' M and P plants.

4 Patenting and the Proximity of MP Firms' M and P Plants

Results in the previous section indicate that firms with both manufacturing (M) and innovation-related (P) establishments patent more than other firms. In this section, we show that among firms with both types of establishments, *i.e.*, MP firms, patents are higher when the minimum distance between M and P facilities is lower. We first document the spatial distribution of M and P establishments among M and P firms over our sample period. We then estimate the relationship between firm patenting and these distances using a series of panel regressions, with and without firm fixed effects.

4.1 M-P Plant Distance

For each MP firm in each year, we compute the distance between all pairs of M and P establishments within the firm using the geocode information described in Section 2.1. The minimum and average distances between M and P establishments within firm f in year t are denoted $dist_{ft}^{min}$ and $dist_{ft}^{avg}$. Table 3 presents the yearly means and medians of these distances.[29] The starkest pattern is the sharp skewness in M-P plant proximity within firms, manifest in the large gap between $dist_{ft}^{min}$ and $dist_{ft}^{avg}$ across years. In 1977, the median firm's *minimum* distance is only 3 miles, while the median firm's *average* distance is 301 miles – 100 times the minimum. A similar skewness is evident when comparing the means of $dist_{ft}^{min}$ and $dist_{ft}^{avg}$. Thus, although MP firms' manufacturing and innovation plants tend to be hundreds of miles apart, at least one pair of these establishments tends to be very close.[30]

Table 3 also depicts rising distances between a firm's M and P plants over time. Columns 1 and 2 show the median $dist_{ft}^{min}$ doubling from 3 to 6 miles and the mean $dist_{ft}^{min}$ increasing about 44 percent, from 95 to 137 miles. $dist_{ft}^{avg}$ also rises, though not as substantially: its median increases about 38 percent, from 301 to 416 miles, while the mean grows about 16 percent, from 445 to 517

[29] Distances are rounded and medians are constructed as the average of all establishment pairs in the 49^{th} through 51^{st} percentiles, per Census disclosure guidelines.

[30] An example of this type of spatial variation is evidence in Pharmaceutical Manufacturing. For example, publicly available data show that the Danish firm Novo Nordisk has 7 manufacturing locations and 4 R&D facilities around the world. Of the 4 R&D facilities, 2 are colocated with production sites: one in Denmark and one in China. These colocated production sites are often used for testing of new products and for production of complex goods that require continual input from scientists. In contrast, production of mature and simpler compounds tends to occur in the low-wage locations. Appendix Figure A.5 shows these locations. Similarly, Appendix Figure A.4 shows the worldwide locations for Bristol Meyers Squibb based on publicly available information.

Table 3: Distances between MP Firm's M and P Establishments

	Minimum ($dist_{ft}^{min}$)		Average ($dist_{ft}^{avg}$)	
	Mean	Median	Mean	Median
1977	95	3	445	301
1982	115	4	457	322
1987	120	5	470	336
1992	141	6	487	359
1997	153	6	502	381
2002	139	5	501	387
2007	142	5	498	383
2012	137	6	517	416

Source: LBD, BR, and authors' calculations. Table reports summary statistics for the minimum and mean distances (in miles) between MP firm f's M and P establishments in each Census year t, denoted $dist_{ft}^{min}$ and $dist_{ft}^{avg}$. The first two columns report the mean and median for $dist_{ft}^{min}$, while the second two columns report the analogous statistics for $dist_{ft}^{avg}$. The set of firms included in the counts is restricted to the regression sample used in Tables 4 and 5.

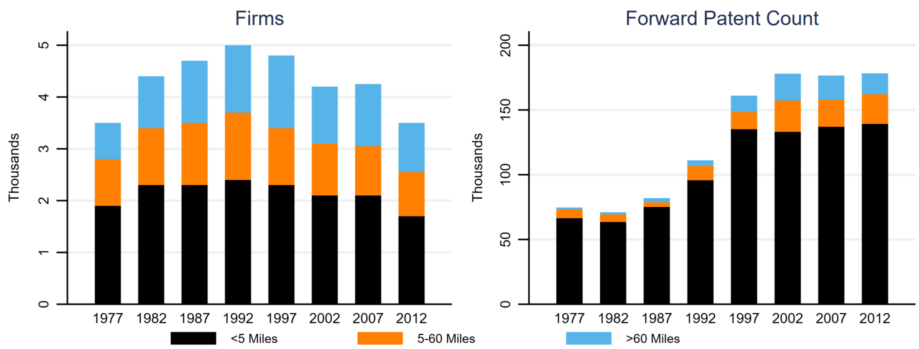
miles. Despite the larger percentage increases for the minimum distances, they remain significantly smaller than average distances, and are easily conducive to repeat visits within a day.

These summary statistics motivate our construction of three dummy variables indicating whether the minimum distance between MP firms' M and P facilities are within 5 miles of each other, $dist_{ft}^{min} \in (0, 5)$, between 5 and 60 miles of each other, $dist_{ft}^{min} \in (5, 60)$, or greater than 60 miles apart, $dist_{ft}^{min} > 60$. The first category captures firms in which the two types of establishments are sufficiently close for employees to travel between them at a low cost, for example such as those within industrial parks, while the second captures locations that are within about an hour's drive.

The left panel of Figure 9 displays the distribution of MP firms across these dummy variables and Census years. As indicated in the figure, MP firms rise from about 3.5 to 5 thousand between 1977 and 1992 before falling to their initial level in 2012. This rise and fall pattern occurs within all three distance categories, though the share of firms with at least one pair of M and P plants less than five miles falls from 54 to 49 percent, while the share of firms with a minimum distance greater than 60 miles rises from 20 to 27 percent.

The right panel of Figure 9 displays firms' patenting activity by their minimum M-P plant distance. As in the previous section, we compute the sum of subsequently granted patents applied for by each firm in years t to $t+4$. We then aggregate these counts across each set of firms. As indicated in the figure, patent counts rise substantially in the 1990s, primarily due to firms whose M and P plants are within 5 miles of each other. This growth, however, is relatively slower than the growth in the forward patent counts of the remaining MP firms, with the result that the share of grants awarded to firms with the closest plants falls from 89 to 78 percent over the sample period. The largest growth in relative terms is the forward patent count of firms whose M and P plants are more than 60 miles

Figure 9: MP Firms and Patents by Minimum M-P Establishment Distance



Source: LBD, BR, and authors' calculations. Left panel displays a breakdown of MP firms across Census years according to the minimum distance ($dist_{ft}^{min}$) between their M and P establishments. Right panel reports the analogous distribution of the sum of these firms' subsequently granted patents applied for in years t to $t+4$. MP firms contain both manufacturing and innovation establishments. P establishments consist of plants in the following (NAICS) sectors: Architectural, Engineering, and Related Services (5413); Specialized Design Services (5414); Computer Systems Design and Related Services (5415); Management, Scientific, and Technical Consulting Services (5416); Scientific Research and Development Services (5417); Software Publishers (5112); Telecommunications (517); Data Processing, Hosting, and Related Services (518); and Corporate, Subsidiary, and Regional Managing Offices (551114). Firm and patent counts are rounded per Census disclosure guidelines. The set of firms included in the counts is restricted to the regression sample used in Tables 4 and 5.

apart. Their share rises from 2 to 9 percent.

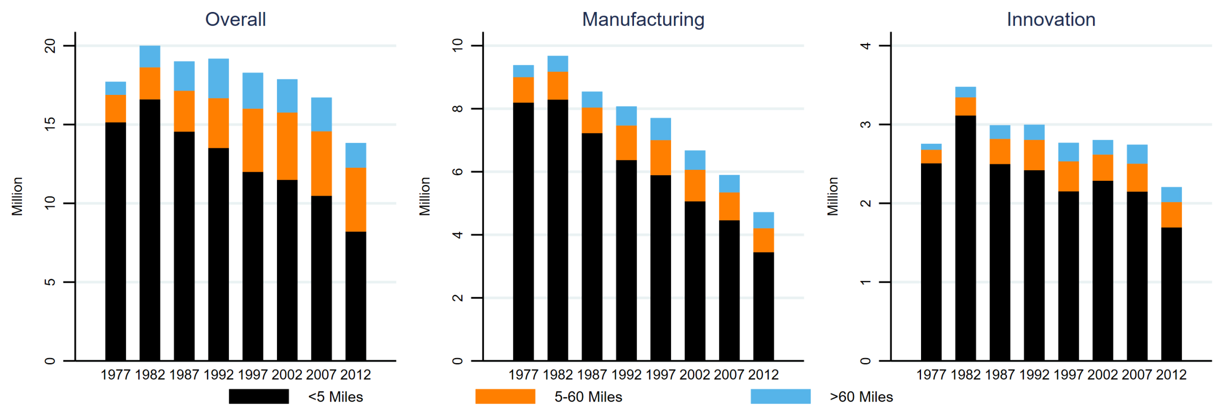
Figure 10 displays analogous distributions of firms' overall as well as M and P employment.[31] As indicated in the figure, the overall employment of firms with the closest M and P establishments declines steadily starting in 1992. Part of this decline is due to the shifting of firms to the more distant bins over time, as noted above, with the result that the employment of firms in the middle distance category increases from 10 to 29 percent of all MP firms' employment, while that of firms in shortest distance category falls from 85 to 59 percent.

The middle panel of Figure 10 reveals that M employment falls more sharply than P employment, particularly among firms whose M-P plants are within 5 miles of each other. Indeed, M employment drops by more than half, from 8.2 to 3.4 million, among these firms. Their innovation employment, reported in the right panel of the figure, also falls substantially, from a high of 3.1 million in 1982 to a low of 1.7 million in 2012.

Figure 11 reports MP firms' average overall, M, and P employment across Census years according to the minimum distance between their M and P plants. As indicated in the first panel, average employment falls substantially among MP firms with the closest M and P plants, from about 8 to less than 5 thousand. By contrast, firms in the middle distance category grow substantially, such that their average employment is about equal to the size of the most colocated firms by 2012. The decline in employment of the MP firms with the closest M and P plants is even more pronounced for manufacturing workers. As a result, these firms reallocate towards P workers over the sample period, but maintain colocation even as their US manufacturing presence shrinks. Average M employment is steady for MP firms whose M and P plants are furthest apart, and growing among MP firms

[31] Firm's overall employment includes workers outside M and P, and is not shown.

Figure 10: Employment of MP Firms by Minimum M-P Establishment Distance



Source: LBD, BR, and authors' calculations. Panels display the distribution of MP firms' total, manufacturing and innovation employment across Census years according to the minimum distance ($dist_{ft}^{min}$) between their M and P establishments. MP firms contain both manufacturing and innovation establishments. P establishments consist of plants in the following (NAICS) sectors: Architectural, Engineering, and Related Services (5413); Specialized Design Services (5414); Computer Systems Design and Related Services (5415); Management, Scientific, and Technical Consulting Services (5416); Scientific Research and Development Services (5417); Software Publishers (5112); Telecommunications (517); Data Processing, Hosting, and Related Services (518); and Corporate, Subsidiary, and Regional Managing Offices (551114). Firm and patent counts are rounded per Census disclosure guidelines. The set of firms included in the counts is restricted to the regression sample used in Tables 4 and 5. Firm's overall employment includes workers outside M and P, and is not shown. This "other" employment is not shown separately.

in the middle distance category. These firms P employment grows more however, so that they also reallocate employment away from manufacturing towards innovation.

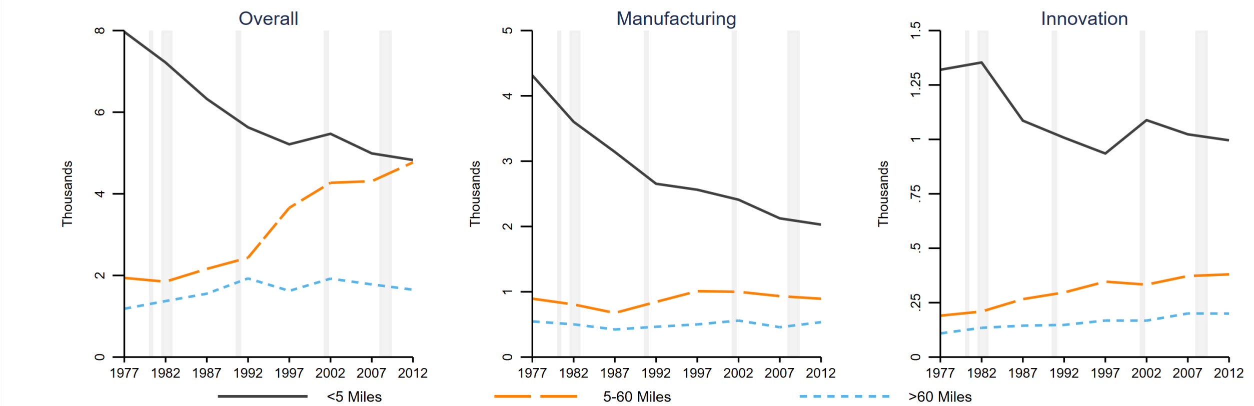
Finally, Figure 12 provides context for the aggregate importance of MP firms. The first panel shows that MP firms represent about 0.1 percent of all firms, and 1.5 percent of manufacturing firms. Despite these small shares, MP firms account for more than half of all US manufacturing employment until the 1990s. By 2012, their share of aggregate manufacturing employment had fallen more than 10 percentage points from its high in 1982, to just above 40 percent. MP firms' share of total employment falls even more dramatically, from about 28 percent in 1977 to just above 10 percent in 2012. The last panel of Figure 12 shows that MP firms patent disproportionately more than other firms. Their share of manufacturing firms' patents is fairly steady at about 80 percent over the entire period, while their share of all firms' patents declines steadily from 70 percent in 1977 to just above 40 percent in 2012. Clearly, further examination of MP firms is necessary for understanding the evolution of US innovation.

4.2 Patenting and M-P Establishment Distance

We investigate the relationship between patenting and M-P plant distance within MP firms using a simple OLS panel specification:

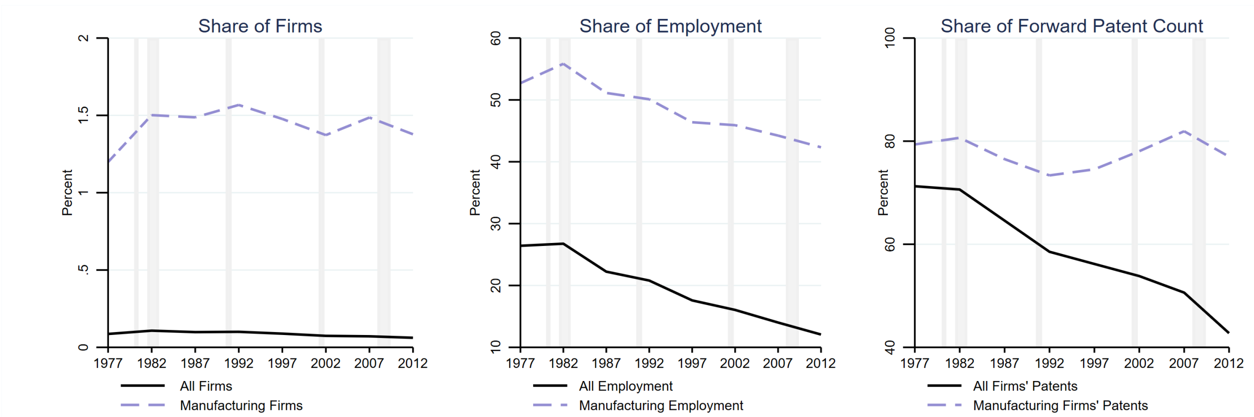
$$\begin{aligned} ln(\widetilde{y}_{ft}) = \quad & \delta_1 \left[dist_{ft}^{min} \in (0,5)\right] + \delta_2 \left[dist_{ft}^{min} \in (5,60)\right] + \\ & \gamma ln(PatentStock_{f,t-1}^{dep}) + \beta X_{ft} + \alpha_t + \alpha_r + \varepsilon_{frt}, \end{aligned} \tag{3}$$

Figure 11: Average Employment per MP Firm by Minimum M-P Establishment Distance



Source: LBD, BR, and authors' calculations. Panels display the weighted average employment of MP firms' total, manufacturing and innovation workers across Census years according to the minimum distance ($dist_{ft}^{min}$) between their M and P establishments. MP firms contain both manufacturing and innovation establishments. P establishments consist of plants in the following (NAICS) sectors: Architectural, Engineering, and Related Services (5413); Specialized Design Services (5414); Computer Systems Design and Related Services (5415); Management, Scientific, and Technical Consulting Services (5416); Scientific Research and Development Services (5417); Software Publishers (5112); Telecommunications (517); Data Processing, Hosting, and Related Services (518); and Corporate, Subsidiary, and Regional Managing Offices (551114). Firm and patent counts are rounded per Census disclosure guidelines. The set of firms included in the counts is restricted to the regression sample used in Tables 4 and 5. This "other" employment is not shown separately.

Figure 12: MP Firms' Share of Firms, Employment, and Patent Grants



Source: LBD, BR, and authors' calculations. Panels display MP firms' share of firms, employment and forward patent grants across Census years. MP firms contain both manufacturing and innovation establishments. P establishments consist of plants in the following (NAICS) sectors: Architectural, Engineering, and Related Services (5413); Specialized Design Services (5414); Computer Systems Design and Related Services (5415); Management, Scientific, and Technical Consulting Services (5416); Scientific Research and Development Services (5417); Software Publishers (5112); Telecommunications (517); Data Processing, Hosting, and Related Services (518); and Corporate, Subsidiary, and Regional Managing Offices (551114). Firm and patent counts are rounded per Census disclosure guidelines. The set of firms included in the counts is restricted to the regression sample used in Tables 4 and 5. This "other" employment is not shown separately.

where $\widetilde{y}_{ft}$ and X_{ft} are defined as in Equation 1.[32] The first two terms on the right-hand-side are the minimum-distance indicator variables defined in the previous section, while $ln(PatentStock^{dep}_{f,t-1})$ represents firm f's one-year-lagged and depreciated stock of patent grants.[33] As in the previous section, we also include full sets of year and major-county fixed effects, and estimate specifications with and without firm fixed effects. The latter are particularly revealing as the relationship between M-P plant distance and patenting is identified solely from variation in these distances within MP firms over time. $dist_{ft} > 60$ is the omitted distance category, so that δ_1 is interpreted as the average log-point difference in the number of patents accumulated by firms with M and P plants within 5 miles of each other, relative to firms that whose plants are more than 60 miles apart. The estimation sample consists of all MP firms in Census years from 1977 to 2012 that are present in at least two EC years.[34] We cluster the standard errors by firm.

Table 4 presents the results from estimating Equation (3) using firms' overall forward patent count as the dependent variable.[35] The results in Column 1 indicate that firms' whose minimum M-P plant distance is less than 5 miles have granted patents that are 13 log points higher than firms whose plants are more than 60 miles apart. Column 2 repeats this specification, but with firm fixed effects, and magnitudes increase: firms with the closest M and P plants have patents that are estimated to be 15 log points higher, while the estimated impact is 9.8 log points for firms in the middle distance category. In Columns 3 and 4 we repeat the specifications and control for the firm's lagged depreciated stock of patent grants. Estimates in the specification with firm fixed effects remain large and statistically significant.

Table 5 presents results from estimating Equation (3) using firms' granted patents' citations, manufacturing patents, and processing patents as dependent variables. We restrict attention to the specifications with firm fixed effects. Results in the first two columns reveal that patent quality, like patents, is higher when firms' M and P plants are more proximate. They also exhibit a similar decay across the two distance categories: firms earn approximately 25 log points more citations when their M and P plants are within 5 miles of each other, relative to when than those whose plants are greater than 60 miles apart. When plants are between 5 and 60 miles apart, the relative citation premium is 13 to 14 log points.

The final four columns of Table 5 examine firms' manufacturing and process patents. Here, too, we find that patents are higher when firms' M and P plants are closer. For manufacturing patents, the coefficient estimates are similar in magnitude and significance to overall patents, which is not unexpected given that manufacturing patents represent the bulk of all patents. Finally, columns 5 and 6 indicate that when firms' M and P plants are less than 5 miles apart, they exhibit processing

[32] In this regression, the firm-size covariates as well as firm fixed effects control for the probability that colocation is higher among firms that have more establishments because they have more establishment pairs that could be flagged as colocated, but the Firm FE and size controls should cover that

[33] We compute these stocks as the discounted sum of firm f granted patent applications from 1977 to year $t-1$, where the discount rate is 0.15, which follows the literature (see Hall (2006) for a discussion on this topic).

[34] This restriction is necessary for the specification with firm fixed effects. We limit the specifications without firm fixed effects to the same sample both to allow for a comparison of the estimates and to facilitate disclosure. The results from estimating Equation (3) on all MP firms are qualitatively the same.

[35] As defined above, a firm's forward patent count is the sum of subsequently granted patents applied for in years t to $t+4$.

Table 4: Patenting Among MP Firms

Dependent variable is: $ln(Patents_{f,t:t+4})$

	(1)	(2)	(3)	(4)
$dist_{ft}^{min} \in (0,5)$	0.131***	0.149***	0.0201	0.116***
	(0.0284)	(0.0300)	(0.0131)	(0.0279)
$dist_{ft}^{min} \in (5,60)$	-0.0230	0.0984***	0.00690	0.0764***
	(0.0303)	(0.0298)	(0.0148)	(0.0281)
$ln(Patent\ Stock_{f,t-1}^{dep})$			0.833***	0.278***
			(0.00526)	(0.0148)
$Workers_{ft}$				
10 - 99	0.0543	0.0121	0.0133	-0.0366
	(0.0513)	(0.0471)	(0.0333)	(0.0509)
100 - 499	0.365***	0.0902*	0.0612*	0.0178
	(0.0517)	(0.0495)	(0.0330)	(0.0529)
500 - 4999	1.273***	0.283***	0.193***	0.172***
	(0.0562)	(0.0544)	(0.0340)	(0.0570)
5000+	3.125***	0.868***	0.504***	0.638***
	(0.0866)	(0.0721)	(0.0405)	(0.0714)
Age_{ft}				
5 - 9	-0.0871*	-0.0710**	-0.136***	-0.0777**
	(0.0453)	(0.0346)	(0.0364)	(0.0337)
10+	-0.108**	-0.115**	-0.236***	-0.130***
	(0.0536)	(0.0490)	(0.0351)	(0.0466)
Year Fixed Effects	Yes	Yes	Yes	Yes
FIPS Fixed Effects	Yes	Yes	Yes	Yes
Firm Fixed Effects	No	Yes	No	Yes
R-Squared	0.401	0.875	0.787	0.881
Observations	34,500	34,500	34,500	34,500

Source: LBD, BR, USPTO and authors' calculations. Table reports the results of estimating Equation 3 on US firms with both M and P establishments in at least two Census years from 1977 to 2012. Dependent variable is the inverse hyperbolic sine transformation of the sum of subsequently granted patents applied for by firm f in years t to $t+4$. Its mean and standard deviation are 1.114 and 1.768. $dist_{ft}^{min} \in (0,5)$ and $dist_{ft}^{min} \in (5,60)$ are dummy variables indicating whether the M and P establishments are within 5 miles of each other, or 5 to 60 miles apart. The omitted category is firms with M and P establishments over 60 miles apart. FIPS fixed effects capture the main FIPS code (i.e., county) for each firm, based on its employment. Standard errors clustered by firm. Number of observations rounded per Census disclosure guidelines.

patent grants that are 10 and 6 log points higher. Coefficient estimates for $dist_{ft}^{min} \in (5, 60)$ are also correspondingly lower, at 6 and 4 log points. This difference is equal to about 12 percent of the average number of processing patents per firm (0.0682/0.588), which is roughly equivalent to the implied impacts for citations and manufacturing patents, and slightly larger than the implied effect for overall patents.[36]

Table 5: Patenting Among MP Firms

Dependent variable is:	$ln(Citations_{f,t:t+4})$		$ln(ManufPatents_{f,t:t+4})$		$ln(ProcessingPatents_{f,t:t+4})$	
	(1)	(2)	(3)	(4)	(5)	(6)
$dist_{ft}^{min} \in (0, 5)$	0.258***	0.243***	0.146***	0.115***	0.101***	0.0682***
	(0.0519)	(0.0514)	(0.0282)	(0.0263)	(0.0223)	(0.0204)
$dist_{ft}^{min} \in (5, 60)$	0.143***	0.133**	0.0929***	0.0721***	0.0634***	0.0415**
	(0.0528)	(0.0523)	(0.0279)	(0.0265)	(0.0223)	(0.0208)
$ln(Patent\ Stock_{f,t-1}^{dep})$		0.126***		0.264***		0.278***
		(0.0228)		(0.0147)		(0.0136)
$Workers_{ft}$						
10 - 99	-0.102	-0.124	0.0141	-0.0321	0.0212	-0.0275
	(0.104)	(0.106)	(0.0440)	(0.0474)	(0.0303)	(0.0335)
100 - 499	0.0301	-0.00277	0.0792*	0.0105	0.0439	-0.0285
	(0.108)	(0.110)	(0.0462)	(0.0492)	(0.0316)	(0.0345)
500 -4999	0.326***	0.276**	0.252***	0.147***	0.131***	0.0207
	(0.117)	(0.118)	(0.0507)	(0.0530)	(0.0346)	(0.0370)
5000+	1.144***	1.040***	0.810***	0.592***	0.592***	0.362***
	(0.138)	(0.138)	(0.0679)	(0.0671)	(0.0539)	(0.0516)
Age_{ft}						
5 - 9	-0.0553	-0.0583	-0.0445	-0.0508	-0.0265	-0.0331
	(0.0759)	(0.0760)	(0.0339)	(0.0326)	(0.0264)	(0.0249)
10+	-0.116	-0.123	-0.0575	-0.0720	-0.0203	-0.0356
	(0.0991)	(0.0990)	(0.0473)	(0.0447)	(0.0386)	(0.0353)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
FIPS Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
R-Squared	0.834	0.835	0.877	0.883	0.860	0.872
Observations	34,500	34,500	34,500	34,500	34,500	34,500
Dep Var Mean & SD	1.955, 2.843		1.017, 1.686		0.588, 1.271	

Source: LBD, BR, USPTO and authors' calculations. Table reports the results of estimating Equation 3 on US firms with both M and P establishments in at least two Census years from 1977 to 2012. First dependent variable is inverse hyperbolic sine transformation of the subsequently granted patents firm f applied for in years t to $t+4$. Remaining dependent variables are the inverse hyperbolic sine transformation of subsequently granted manufacturing and processing patents applied for by firm f in years t to $t+4$. $dist_{ft}^{min} \in (0, 5)$ and $dist_{ft}^{min} \in (5, 60)$ are dummy variables indicating whether the M and P establishments are within 5 miles of each other, or 5 to 60 miles apart. The omitted category is firms with M and P establishments over 60 miles apart. FIPS fixed effects capture the main FIPS code (i.e., county) for each firm, based on its employment. Standard errors clustered by firm. Number of observations rounded per Census disclosure guidelines.

The results in this section point to a strong relationship between the proximity of a firm's M and

[36]In Table A.1 of Appendix Section E we present analogous regression results for specification without firm fixed effects. In those results, the estimated coefficient for $dist_{ft}^{min} \in (0, 5)$ is positive and statistically significant in five of the six specifications.

P plants and their patenting output. Clearly, however, plant location is an endogenous decision of the firm. Plant location is also affected by macroeconomic factors that may separately influence a firm's ability to innovate. For example, a firm may shutter manufacturing plants due to increased import competition, and this import competition may also reduce the firm's size and thus its ability to invest in R&D.

To address these possibilities and investigate the impact of colocation on innovation, we plan to analyze where firm innovation occurs, as described in the next section. We also aim to analyze the source of variation in $dist_{ft}^{min}$ within firms, *i.e.*, whether it is driven by M or P plant entry or exit, or changes in the major activity of the plant. Exploring these margins will help tease apart the different channels by which the results documented in this section are generated.

5 Firm-Region-Level Regressions

The firm-level regressions presented in the last section establish a link between patenting and firms that have colocated manufacturing and innovation facilities. In a future draft, we will exploit the firm-city-state patent data to investigate whether the patenting by these firms occurs in the regions where those establishments are colocated using a simple panel regression

$$\begin{aligned} ln(\widetilde{y}_{frt}) = \quad & \gamma_1 M_{frt} + \gamma_2 M_{frt} + \gamma_3 M_{ft} * P_{ft} + \\ & \beta X_{ft} + \alpha_f + \alpha_t + \alpha_r + \varepsilon_{frt}. \end{aligned} \tag{4}$$

6 Conclusion

To be written.

References

Aghion, Philippe, Nick Bloom, Richard Blundell, Rachel Griffith, and Peter Howitt, "Competition and Innovation: An Inverted U-Relationship," *The Quarterly Journal of Economics*, 2005, *120* (2), 701–728.

Arkolakis, Costas, Natalia Ramondo, Andrès Rodrìgues-Clare, and Stephen Yeaple, "Innovation and Production in the Global Economy," *American Economic Review*, 2018, *108* (8), 2128–2173.

Audretsch, David B. and Maryann P. Feldman, "R&D Spillovers and the Geography of Innovation and Production," *American Economic Review*, 1996, *86* (3), 630–40.

Autor, David, David Dorn, Gordon H Hanson, Gary Pisano, and Pian Shu, "Foreign Competition and Domestic Innovation: Evidence from US Patents," *American Economic Review: Insights*, forthcoming.

Autor, David H., David Dorn, and Gordon H. Hanson, "The China Syndrome: Local Labor Market Effects of Import Competition," *American Economic Review*, 2013, *103* (6), 2121–2168.

Balasubramanian, Natarajan and Jagadeesh Sivadasan, "What happens when firms patent? New evidence from US economic census data," *The Review of Economics and Statistics*, 2011, *93* (1), 126–146.

Berkes, Enrico and Ruben Gaetani, "The Geography of Unconventional Innovation," mimeo, University of Toronto 2020.

Bernard, Andrew B. and Teresa C. Fort, "Factoryless Goods Producing Firms," *American Economic Review: Papers & Proceedings*, 2015, *105* (5), 518–23.

_ , J. Bradford Jensen, and Peter K. Schott, "Survival of the Best Fit: Exposure to Low-Wage Countries and the (Uneven) Growth of U.S. Manufacturing Plants," *Journal of International Economics*, 2006, *68*, 219–237.

_ , _ , Stephen J. Redding, and Peter K. Schott, "Firms in International Trade," *Journal of Economic Perspectives*, 2007, *21* (3), 105–130.

_ , _ , _ , and _ , "Wholesalers and Retailers in US Trade," *American Economic Review*, May 2010, *100* (2), 408–13.

Bernard, Andrew B, Teresa C Fort, Valerie Smeets, and Frederic Warzynski, "Heterogeneous Globalization: Offshoring and Reorganization," Working Paper 26854, National Bureau of Economic Research March 2020.

Bernard, Andrew B., Valerie Smeets, and Frederic Warzynski, "Rethinking Deindustrialization," *Economic Policy*, 2017, *32* (89), 5–38.

Bilir, L. Kamran and Eduardo Morales, "Innovation in the Global Firm," *Journal of Political Economy*, 2020, *124* (4), 1566–1625.

Bloom, Nicholas, Charles I. Jones, John Van Reenen, and Michael Webb, "Are Ideas Getting Harder to Find?," *American Economic Review*, 2020, *110* (4), 1104–1144.

_ , Mirko Draca, and John Van Reenen, "Trade Induced Technical Change: The Impact of Chinese Imports on Innovation, Diffusion, and Productivity," *Review of Economic Studies*, 2016, *83*, 87–117.

Bloom, Nick, Kyle Handley, Andre Kurman, and Philip Luck, "The Impact of Chinese Trade on US Employment: The Good, the Bad, and the Debatable," mimeo, Stanford University 2019.

Boler, Esther Ann, Andreas Moxnes, and Karen Helene Ulltveit-Moe, "R&D, International Sourcing and the Joint Impact on Firm Perform," *American Economic Review*, 2015, *105*, 3704–3739.

Branstetter, Lee, Jong-Rong Chen, Britta Glennon, and Nikolas Zolas, "Does offshoring production harm innovation? Firm-level evidence from Taiwan," Working paper 2020.

Buzard, Kristy, Gerald A. Carlino, Robert M. Hunt, Jake K. Carr, and Tony E. Smith, "The Agglomeration of American R&D Labs," *Journal of Urban Economics*, 2017, *101*, 14–26.

_ , _ , _ , _ , and _ , "Localized Knowledge Spillovers: Evidence from Spatial Clustering of R&D Labs and Patent Citations," *Regional Science and Urban Economics*, forthcoming.

Davis, Donald R. and Jonathan I. Dingel, "A Spatial Knowledge Economy," *American Economic Review*, 2019, *109* (1), 153–70.

Delgado, Mercedes, "The Co-Location of Innovation and Production in Clusters," *Industry and Innovation*, 2020, pp. 842–70.

Ding, Xiang, Teresa C Fort, Stephen J Redding, and Peter K Schott, "Structural change within versus across firms: Evidence from the United States," mimeo, Dartmouth College 2019.

Dreisigmeyer, David, Nathan Goldschlag, Marina Krylova, Wei Ouyang, Elisabeth Perlman et al., "Building a Better Bridge: Improving Patent Assignee-Firm Links," Technical Report, Center for Economic Studies, US Census Bureau 2018.

Duranton, Giles and Henry G. Overman, "Testing for Localization Using Micro-Geographic Data," *Review of Economic Studies*, 2005.

Ellison, Glenn and Edward L. Glaeser, "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach," *Journal of Political Economy*, 1997.

_ , _ , and William R. Kerr, "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns," *American Economic Review*, 2010, *100*, 1195–1213.

Fort, Teresa C., "Technology and Production Fragmentation: Domestic versus Foreign Sourcing," *Review of Economic Studies*, 2017, *84* (2), 650–687.

_ and Shawn D. Klimek, "The Effects of Industry Classification Changes on US Employment Composition," Working Paper 18-28, Center for Economic Studies 2018.

_ , Justin R. Pierce, and Peter K. Schott, "New Perspectives on the Decline of US Manufacturing Employment," *Journal of Economic Perspectives*, 2018, *32* (2), 47–72.

Fort, Teresa C, Peter K Schott, and Nikolas Zolas, "Measuring US Innovation," mimeo, US Census Bureau 2020.

Fuchs, Erica and Randolph Kirchain, "Design for Location? The Impact of Manufacturing Offshore on Technology Competitiveness in the Optoelectronics Industry," *Management Science*, 2010, *56* (12), 2323–2349.

Ganglmair, Bernhard, W. Keith Robinson, and Michael Seeligson, "Patent Claims: New Data and Stylized Facts," *Mimeo*, 2020.

Goldschlag, Nathan, Travis J Lybbert, and Nikolas J Zolas, "Tracking the technological composition of industries with algorithmic patent concordances," *Economics of Innovation and New Technology*, 2019, pp. 1–21.

Graham, Stuart JH, Cheryl Grim, Tariqul Islam, Alan C Marco, and Javier Miranda, "Business dynamics of innovating firms: Linking US patents with administrative data on workers and firms," *Journal of Economics & Management Strategy*, 2018, *27* (3), 372–402.

Greenland, Andrew N, Mihai Ion, John W Lopresti, and Peter K Schott, "Using Equity Market Reactions to Infer Exposure to Trade Liberalization," Working Paper 27510, National Bureau of Economic Research July 2020.

Griliches, Zvi, "Productivity, R&D, and the Data Constraint," *The American Economic Review*, 1994, *84* (1), 1–23.

Hall, Bronwyn, "R&D, productivity and market value," Working Paper, IFS 2006.

Haltiwanger, John, Ron S. Jarmin, and Javier Miranda, "Who Creates Jobs? Small Versus Large Versus Young," *The Review of Economics and Statistics*, 2013, *95*, 347–361.

Jaffe, Adam B., Manuel Trajtenberg, and Rebecca Henderson, "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," *Quarterly Journal of Economics*, 1993, *108* (3), 577–598.

Jarmin, Ron S. and Javier Miranda, "The Longitudinal Business Database," CES Working Paper 02-17 2002.

Kamal, Fariha, "A Portrait of US Factoryless Goods Producers," Working Paper 25193, NBER 2018.

_ **and Wei Ouyang**, "Identifying US Merchandise Traders: Integrating Customs Transactions with Business Administrative Data," CES Working Paper 20-28 2020.

Kerr, William R and Shihe Fu, "The survey of industrial R&D—patent database link project," *The Journal of Technology Transfer*, 2008, *33* (2), 173–186.

Kortum, Samuel S., "Research, Patenting, and Technological Change," *Econometrica*, 1997, *65* (6), 1389–1419.

Kota, Sridhar and Tom Mahoney, "Innovation Should Be Made in the U.S.A.," 2019.

Lan, Ting, "The Coagglomeration of Innovation and Production," mimeo, University of Michigan 2019.

Lybbert, Travis J and Nikolas J Zolas, "Getting patents and economic data to speak to each other: An 'algorithmic links with probabilities' approach for joint analyses of patenting and economic activity," *Research Policy*, 2014, *43* (3), 530–542.

Naghavi, Alireza and Gianmarco Ottaviano, "Offshoring and Product Innovation," *Economic Theory*, 2009, *38*, 517–32.

Pierce, Justin R. and Peter K. Schott, "The Surprisingly Swift Decline of U.S. Manufacturing Employment," *American Economic Review*, 2016, *106* (7), 1632–1662.

Pisano, Gary P. and Willy C. Shih, "Does America Really Need Manufacturing?," *Harvard Business Review*, 2012.

Rodríguez-Clare, Andrés, "Offshoring in a Ricardian World," *American Economic Journal: Macroeconomics*, April 2010, *2* (2).

Tecu, Isabel, "The Location of Industrial Innovation: Does Manufacturing Matter?," Working Paper 13-09, Center for Economic Studies 2013.

Trajtenberg, Manuel, *Economic analysis of product innovation: The case of CT scanners*, Vol. 160, Harvard University Press, 1990.

Wuchty, Stefan, Benjamin F Jones, and Brian Uzzi, "The increasing dominance of teams in production of knowledge," *Science*, 2007, *316* (5827), 1036–1039.

Online Appendix - Not For Publication

A Assigning Geocodes to Establishments

This section outlines our procedure for assigning a time-invariant geocode (*i.e.*, a latitude and longitude) to every US establishment appearing in our sample.

A.1 Geocodes from the Business Register

Beginning in 2008, the Business Register provides geocodes for establishments based on their Census block code. We use these geocodes for all plants present in 2008 and subsequent years. We also use them to backfill geocode information for establishments in earlier years that do not exit before 2008.

Three issues arise with Census block geocodes. The first is that all establishments within a Census block are assigned the same (usually centroid) latitude and longitude of that block. Thus, any plants in the same Census block have zero distance to each other. The second issue with Census blocks is that their boundaries can change over time, slightly shifting their latitudes and longitudes. In such cases, we take the modal latitude and longitude across years. The final issue with Census blocks is that they are constructed for estimating the US population and therefore focus on residential areas. Strictly commercial areas are not assigned Census blocks. As a result, establishments in those areas – roughly 20 percent of plants – do not have a geocode in the BR.

Overall, the BR provides us with geocodes for about 30 percent of the establishments in our 1977 to 2016 sample. This rate is around 80 percent for 2008 and onwards, and declines steadily each year removed from 2008, to approximately 10 percent by 1976.

We discuss our procedure for assigning geocodes to the remaining establishments in the next three subsections.

A.2 Geocodes from ArcGIS

If an establishment cannot be matched to a geocode in the post-2007 BR, we first try to manually assign one using its street address, which is also tracked in BR. This address can be either its physical address (collected from the Economic Census) or its mailing address (if different). In preparation for geocode matching, we clean the addresses in the BR by removing non-alphanumeric and superfluous terms such as SUITE, FLOOR, PO BOX, ROOM, APARTMENT and UNIT, and standardizing frequently used terms such as SAINT or NORTH.

We feed this cleaned address information into ArcGIS to recover geocodes. As we are unable make use of online tools with Census address information, we rely on ArcGIS' StreetMap USA for this mapping. One caveat with this approach is that StreetMap USA is a static map, last updated in 2005. Given computational constraints related to running ArcGIS internally, we recover geocodes in the following order. First, we focus on the establishments that are most relevant for our study, namely M and P plants (*i.e.*, NAICS codes beginning with 3, 51, 54 and 55). We then perform ArcGIS searches for all establishments that are part of multi-unit firms. We have not yet recovered geocodes for establishments of single-unit firms.

ArcGIS yields geocodes for 65 to 70 percent of the addresses we enter, with the result that we recover geocodes for another 10 percent of the establishments in our 1977 to 2016 sample. This rate is mostly steady from 1976 to 2000, but declines as the year approaches 2008 (where the majority of establishments are assigned the BR geocode). Together with the geocodes from the BR, we have now matched 40 percent of the establishments in our sample.

A.3 Geocodes from BR Census Tracts

For establishments not yet matched with a geocode, we rely on existing Census Tract, zip code and county crosswalks provided by Census.

Establishments' Census Tracts are available in the BR starting in 1992. For establishments present in the BR after 1991, we assign the latitude and longitude of the tracts' centroid using Census' Census Tract to Geocode crosswalk.[37] From 1992 onwards, this Census tract crosswalk provides more than two-thirds of the non-BR geocodes.

Census Tracts provide geocodes for 30 percent of the establishments in our 1977 to 2016 sample. Together with the geocodes from the BR and ArcGIS, we have now matched 70 percent of the establishments in our sample.

A.4 Geocodes from LBD Zip Codes

We use information on establishments' zip codes from the LBS for the remaining establishments. While these zip codes are drawn initially from the physical address of the BR, they are further filtered and cleaned for use in constructing Census' County Business Patterns (CBP) microdata. As zip code boundaries may also change over time, we construct time-invariant zip code centroid geocodes by taking their mean latitude and longitude across years 1990, 2000, 2010 and 2018. The remaining 30 percent of establishments' geocodes are assigned using this method. This method is also the dominant method for assigning geocodes to establishments born prior to 1992.

A.5 Validation using LBD Zip Codes

It is important to note that we assign the geocodes based on the most precise location information available, subject to computational constraints. If an establishment has a physical address, we use that, but if it is missing and mailing address is available, we use that. We preserve this ranking within each level of our assignment algorithm, but not across them. That is, we prioritize physical addresses over mailing address both in selecting BR geocodes and in selecting ArcGIS geocodes. However, we use a mailing address geocode from the BR over a physical address geocode from ArcGIS.

Having said that, we verify that all of the geocodes lie within the contiguous US, and use the LBD zip codes to check the assignments using geocodes from the BR, ArcGIS and Census Tracts. If the absolute value of the difference between the zip code geocode and the original assignment is greater than 0.5 degrees (approximately 35 miles in latitude and 27 miles in longitude), we replace it with the zip code assignment. This last step may happen for new firms in non-Census years, when the physical address is not yet known by Census but the mailing address may be present in the BR. In this scenario, we would rely on the LBD zip code if the rule above were tripped. We replace approximately 5 percent of our assignments this way.

B Corrections to the LBD's Firm Identifier ($firmid$)

To be completed.

C USPTO PatentView Data

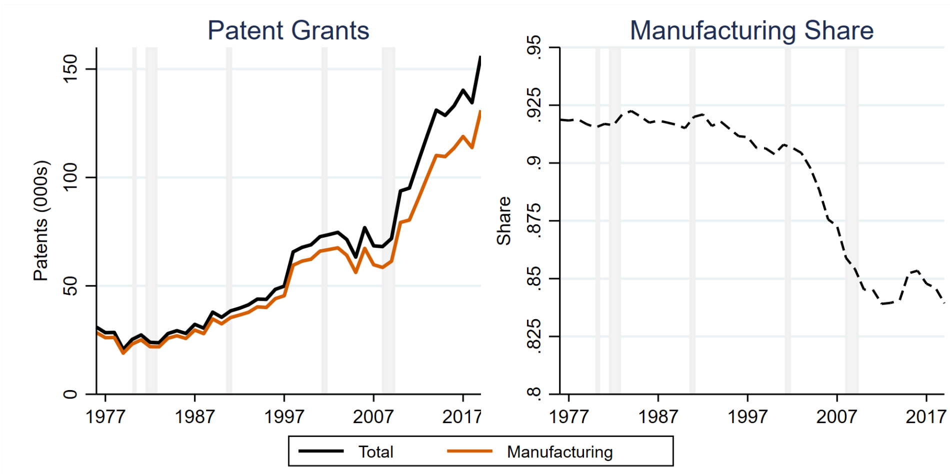
This section provides a brief overview of publicly available US patent data as well as a description of how we match these data to US firms in the Census Bureau's Longitudinal Business Database (LBD).

[37] These crosswalks are available at `https://www.census.gov/geographies/reference-files/time-series/geo/gazetteer-files.html`. We used the 2018 Census tract file based on the 2010 Census.

C.1 USPTO PatentView Data

The left panel of Figure A.1 plots the number of US patent grants in the publicly available USPV data available from the USPTO, overall and for manufacturing. The right panel reports manufacturing patents as a share of overall patents. As indicated in the figure, manufacturing patents account for the largest share of patents, though the growth of these patents slows *vis-à- vis* all other sectors starting in the early 1990s. As a result, manufacturing patents as a share of total patents falls over the sample period, from 92 percent in 1976 to 84 percent in 2017.

Figure A.1: US Patent Grants



Source: USPV and authors' calculations. Left panel plots the number of US patent grants, overall and for manufacturing, by grant year and NAICS sector from 1976 to 2016. Patent counts are on a log scale.

Within manufacturing, we find that patent grants increase in all 3-digit NAICS manufacturing industries except Textiles (NAICS 314) and Printing (NAICS 323) between 1976 and 2016. As reported in the left panel of Figure A.2, growth of manufacturing patents between these years is dominated by Computers (NAICS 334), which increased from 17 to 34 percent (about 45 million patents). This sector includes computers and peripherals, semiconductors, and navigational, measuring, electromedical and control instruments. As noted in the introduction, we find that FMFs are disproportionately rooted in this sector.

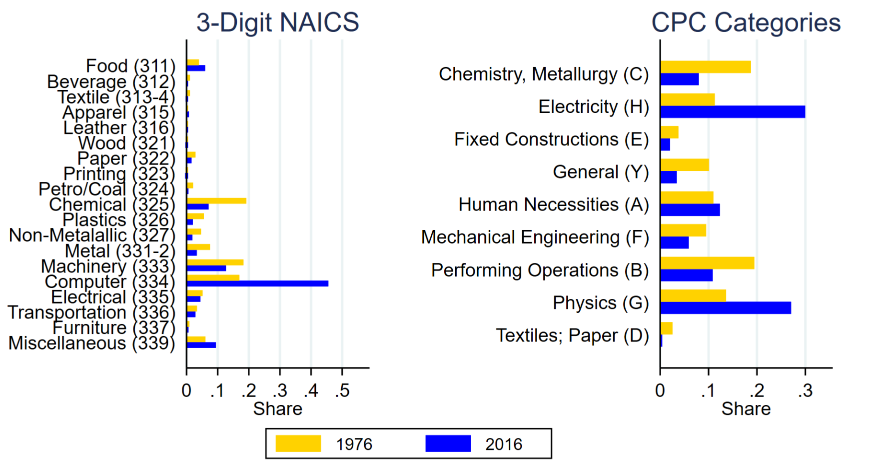
The right panel of Figure A.2 reports a similar breakdown of *overall* US patents by their CPC category and grant year. As illustrated in the figure, by this categorization scheme, patent growth between 1976 and 2016 is particularly strong in Electricity and Physics, two CPC categories that map disproportionately into NAICS 334.

C.2 Matching PatentView Data to US Firms

Our matching procedure builds upon Fort et al. (2020) and prior efforts to construct bridges between USPTO patent data and Census microdata by Kerr and Fu (2008), Balasubramanian and Sivadasan (2011), Graham et al. (2018) and Dreisigmeyer et al. (2018). The former two link firms to patents in the NBER Patent Database via the names and addresses of assignees. This approach, while effective in identifying potential matches, has difficulty in disambiguating and deduplicating multiple matches due to the relatively broad location information, *i.e.*, city and state.[38] Graham et al. (2018) and Dreisigmeyer et al. (2018), by contrast, address this issue by employing a "triangulation" method

[38] "Deduplication" of the matches occurs when a patent and the assignee yield more than one firm match. In many such cases, the firm names are similar, but they contain multiple firm identifiers.

Figure A.2: Distribution of Patent Grants Across Manufacturing Patent Grants by 3-Digit NAICS Sector



Source: USPV and authors' calculations. Left panel plots the distribution of US manufacturing patent grants by patent grant year across 3-digit NAICS manufacturing sectors. Right panel reports breakdown of overall US patents by CPC category.

that incorporates information about inventors from the employee-employer linkages in Census' Longitudinal Employer-Household Dynamics (LEHD) dataset. This method results in higher-quality links but in practice is applicable only to years after 2000, when the employee-employer data are available for a large number of states. While combining these approaches, that is, using the first method until 2000 and the triangulation strategy afterwards, can create an awkward discontinuity.

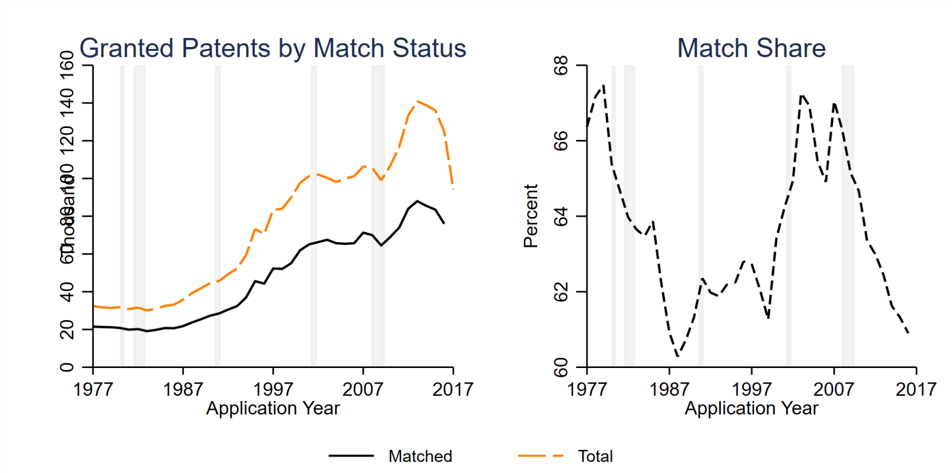
Our approach borrows from both methods. We expand the pool of patent grants to those contained in the USPV, which currently stretch from 1976 to 2020, and use both assignee and inventor name and address information in searching for links, while relying on the tools developed by Kerr and Fu (2008), Balasubramanian and Sivadasan (2011) to assist in the deduplication. This strategy yields a consistent and relatively high match rate of 60 to 67 percent across the sample period for the full set of patent grants assigned to US-based companies between 1976 and 2020. The details of the approach are available in Fort et al. (2020).

The left panel of Figure A.3 plots the total level of granted patents by application year in the USPTO data for patents with assignee type equal to "2 - US Company or Corporation" against our matched sample. The often multiple-year lag between patent applications and grants noted above creates the notable drop in both series towards the end of the sample period: as that end draws near, fewer of the patents applied for in those years have yet been granted. The right panel of Figure A.3 depicts the share of matched patents by year. Shares range from 60 to 67 percent over the sample period. This crosswalk provides the backbone of our analysis into both the levels and types of innovation outcomes that may result from colocation.

To summarize, our matching procedure extends prior attempts to link USPTO patent data to Census microdata by pushing the set of available patents back in time and enlarging the scope of available patents to be linked. Our method yields a consistently high match rate throughout our time period (1977 to 2016) that compares favorably with the prior crosswalks in terms of match rate.[39]

[39] For example, we are able to match nearly twice as many patents as (Kerr and Fu, 2008).

Figure A.3: Matched vs Total Patent Grants by Application Year



USPV, BR, LBD, and authors' calculations. Figure plots the level and shares of US patents for which assignee type is "2 - US Company or Corporation" by application year.

C.3 Matching PatentView Data to Firm-City-States

In this section we outline our procedure for linking USPV patent grants via their inventors to the nearest city and state in which the firm granted the patent has an establishment.[40] As inventors' cities and states often differ from those in which their patents' firms have establishments, we perform this match by calculating the minimum distance from each inventor to each establishment.

We start with our combined patent-to-firm location file and collapse the data to the patent by firm by inventor city by inventor state level, tabulating the number of inventors within each location. We then attempt to assign a geocode to each inventor city and state. Unfortunately, no crosswalk exists that is able to assign a geocode to every city or town in the United States, so we rely on the following approach.

C.3.1 Using the BR City and State Geocodes

Our first step to assigning geocodes to inventors' cities and states is to take all of the Census block-level geocodes listed for a city-state in the BR and take the combined mean geocode across all establishments in that city-state. The benefit of this approach is that it is "economically weighted" by the set of business establishments located within the city and covers most of the city-states in the United States where businesses are located. This information is insufficient for our purposes partly because, as noted in the previous section, geocodes from the BR are based on block-level assignments built primarily to cover residential areas. As a result, commercial zones may be excluded.

C.3.2 Bringing in outside crosswalks

For inventor cities or tows that do not appear in the BR, we incorporate two separate city-state crosswalks: a town-state crosswalk with geocodes provided by the US geological survey (USGS)[41] and a two-stage crosswalk that first assigns the city-state to a zip code, and then assigns a geocode based on the zip code centroid. The USGS crosswalk consists of a "populated place" variable and

[40] We find that both the number of inventors per patent Wuchty et al. (2007) and the number of inventor locations per patent rise over time. For example, the share of inventors based in the United States among patents assigned to US-based firms declines from 91.6 percent in 1977 to 83.3 percent in 2016.

[41] See `https://www.usgs.gov/core-science-systems/ngp/board-on-geographic-names/download-gnis-data`.

state variable. The former are governed by the US Board of Geographic Names (BGN), a federal body created to maintain uniformity in geographic name usage throughout the federal government.

The city-state to zip code crosswalk originates from the social security administration (SSA)[42] Because a city-state often comprises more than one zip code, we randomly select a single zip code to be representative of that city-state. To ensure consistency, we perform a similar geocode assignment to the firm-city-state data coming from the BR and LBD.

Once a city-state has a unique zip code assigned to it from the SSA data, we merge it directly to the inventor city-state before we layer the zip-to-geocode crosswalk. This completes the second step of the geocode assignment for the patent locations.

C.3.3 Google Maps API

As with all administrative data, not all city and state information for US-based inventors can be identified from the BR city-state geocodes, or the SSA zip codes. This is either attributed to misspelling of the city or the city being too small or rural to be counted in either the BR or SSA data. To address this contingency, we take the remaining inventor locations and utilize the Google Maps API and run geocode queries using the given spelling of the city and state. This is our last step in geocoding inventor locations.

The above procedure allows us to assign a unique geocode to more than 99% of the cities and states for US-based USPV inventors. Unfortunately, we are not able to able to assign a geocode for *every* inventor in the US, meaning that there will be some slippage in the patent counts when we assign the patents to exact geographies.

C.3.4 Geocoding Firms' City-States

Once the patent inventor city and states have been assigned a geocode, we perform the exact same steps to assign the geocodes to the city-states in which each firm has an establishment. This firm-city-state database combines the LBD (which has a consistent firm identifier) with the city and state information from the BR. We begin by taking every combination of firm city-states and year. We repeat the steps above in terms of first assigning each city-state a geocode based on the mean establishment geocodes within the BR, and then using the SSA zip code and zip code to geocode crosswalks. For this match, we do not use the Google Maps API.

C.3.5 Combining the patent and firm-city-state

With the above information in hand, we assign the closest firm-city-state to every inventor within our patent-firm data. If there is no direct match, we perform a 1-to-many merge of the firm by city-state by year dataset into the inventor dataset and retain the firm-city-states with the minimum distance, breaking ties arbitrarily.

This step completes our assignment of inventor locations to the nearest city-state in which the firm granted their patent has an establishment.

D M and P Location Examples from Public Data

Figure A.4 displays the locations of Bristol Myers Squibb's (BMS) worldwide establishments as of November 7, 2020. Facilities are grouped into three types: manufacturing plants (MAN), research and development establishments ($R\&D$) and headquarters locations (HQ). Figure A.5 provides similar information for Novo Nordisk. Figure A.6 displays the location of 10 of the 19 inventors listed for BMS patent 10167343. The remaining inventors are in Indiana, California and China.

[42]See `https://www.ssa.gov/policy/docs/statcomps/oasdi_zip/index.html`.

Figure A.4: Bristol Myers Squibb Worldwide Locations

Source: Bristol Myers Squibb, Google Maps and authors' calculations. Map displays Bristol Myers Squibb locations as noted on the firm's website – `https://www.bms.com/about-us/our-company/worldwide-facilities.html` – as of November 7, 2020.

Figure A.5: Novo Nordisk Locations

Source: Novo Nordisk. Figure reports the fifth slide from Novo Nordisk presentation materials posted at `https://en.ppt-online.org/76178` as of November 7, 2020.

E Additional Tables and Figures

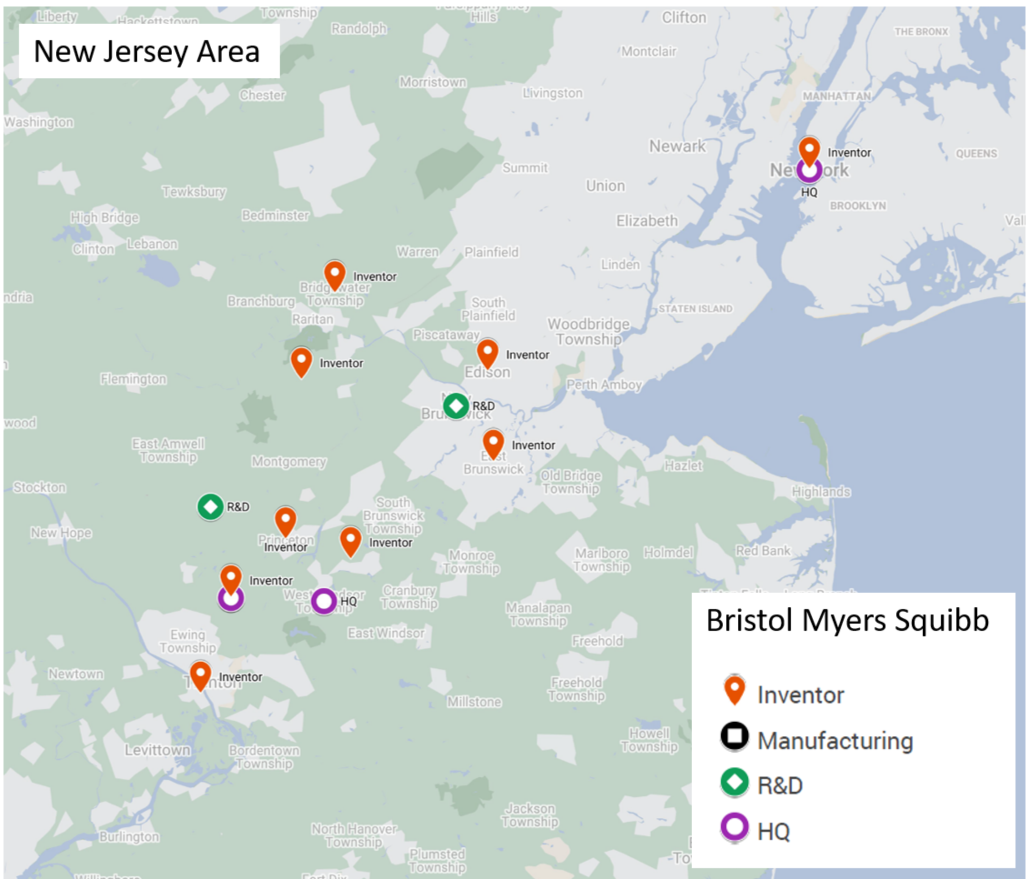
Table A.1 presents results analogous to those in Table 5 of the main text but without firm fixed effects.

Table A.1: Patenting Among MP Firms

Dependent variable is:	$ln(Citations_{f,t:t+4})$		$ln(ManufPatents_{f,t:t+4})$		$ln(ProcessingPatents_{f,t:t+4})$	
	(1)	(2)	(3)	(4)	(5)	(6)
$d^{min}_{ft} \in (0,5)$	0.174***	0.0146	0.143***	0.0376***	0.119***	0.0438***
	(0.0462)	(0.0258)	(0.0270)	(0.0131)	(0.0200)	(0.0123)
$d^{min}_{ft} \in (5,60)$	-0.0477	-0.00459	-0.0164	0.0122	-0.0158	0.00439
	(0.0504)	(0.0297)	(0.0287)	(0.0145)	(0.0206)	(0.0131)
$ln(Patent\ Stock_{f,t-1})$		1.201***		0.795***		0.563***
		(0.0078)		(0.0058)		(0.0082)
$Size2_{ft}$	0.181*	0.121	0.0434	0.0042	0.0010	-0.0268
	(0.0990)	(0.0753)	(0.0496)	(0.0321)	(0.0372)	(0.0260)
$Size3_{ft}$	0.791***	0.353***	0.322***	0.0325	0.131***	-0.0741***
	(0.0998)	(0.0745)	(0.0500)	(0.0318)	(0.0374)	(0.0261)
$Size4_{ft}$	2.347***	0.788***	1.158***	0.127***	0.600***	-0.131***
	(0.105)	(0.0758)	(0.0543)	(0.0330)	(0.0408)	(0.0281)
$Size5_{ft}$	4.986***	1.205***	2.923***	0.423***	2.038***	0.265***
	(0.134)	(0.0851)	(0.0851)	(0.0395)	(0.0706)	(0.0359)
$Age2_{ft}$	-0.102	-0.172**	-0.0803*	-0.127***	-0.0565*	-0.0892***
	(0.0929)	(0.0818)	(0.0426)	(0.0344)	(0.0299)	(0.0260)
$Age3_{ft}$	-0.288***	-0.474***	-0.102**	-0.224***	-0.0401	-0.127***
	(0.104)	(0.0777)	(0.0502)	(0.0331)	(0.0348)	(0.0271)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
FIPS Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
R-Squared	0.385	0.695	0.393	0.780	0.367	0.708
Observations	34,500	34,500	34,500	34,500	34,500	34,500
Dep Var Mean & SD	1.955, 2.843		1.017, 1.686		0.588, 1.271	

Source: LBD, BR, USPTO and authors' calculations. Table reports the results of estimating Equation 3 on US firms with both M and P establishments in at least two Census years from 1977 to 2012. Dependent variables are the inverse hyperbolic sine transformation of total citations, manufacturing patents, and processing patents for patents granted to firm f between Census year t and $t+4$. $dist^{min}_{ft} \in (0,5)$ and $dist^{min}_{ft} \in (5,60)$ are dummy variables indicating whether the M and P establishments are within 5 miles of each other, or between 5 and 60 miles of each other. The omitted category is firms with M and P establishments over 60 miles apart. FIPS fixed effects capture the main FIPS code for a firm, based on its employment. Standard errors clustered by firm. We control for firm employment size and age categories. Number of observations rounded per Census disclosure guidelines.

Figure A.6: Bristol Myers Squibb New Jersey, New York, and Pennsylvania Locations



Source: Bristol Myers Squibb, USPV, Google Maps and authors' calculations. Map displays location of 9 of the 13 inventors associated with Bristol Myers Squibb Patent 10167343. BMS locations are as noted on the firm's website – `https://www.bms.com/about-us/our-company/worldwide-facilities.html` – as of November 7, 2020. Inventor addresses are from USPV.