

# Distance (Still) Hampers Diffusion of Innovations\*

Georg von Graevenitz<sup>a†</sup>

*a Queen Mary University,  
CCP and CREATE*

Stuart J.H. Graham<sup>b</sup>

*b Georgia Institute  
of Technology*

Amanda Myers<sup>c‡</sup>

*c United States Patent &  
Trademark Office.*

August 20, 2019

## ABSTRACT

This paper introduces a new innovation data source to re-examine how spatial distance affects the diffusion of ideas and innovations in an economy. We exploit the descriptions of products and services contained in U.S. trademark registrations during 1980-2012 to identify terms (tokens) not previously used by firms to describe products and services. From these we select tokens frequently re-used by follower firms. By linking the new tokens to the business addresses of innovator and follower firms, our data encompass all instances in which innovations captured by trademark tokens arise within and diffuse across the United States. We aggregate innovations at the year and ZIP code level and estimate Poisson models of the likelihood and intensity of diffusion between locations. After endogenising the creation of new diffusion links between ZIP codes, our results show that spatial distance no longer affects the creation of diffusion links within the US after 1996. However, contingent on previous diffusion from a sending to a receiving ZIP code, we find persistent, strong and negative effects of greater spatial distance on the intensity (extent) of diffusion for existing transfer links between locations within the US.

**KEYWORDS:** Innovation, Diffusion, Rate of Diffusion, Distance, Innovation Index, Trademarks, Patents

**JEL Classification:** O3, O51, R1, R32

---

\*We would like to thank Robert Burrell, Amy Cotton, Carsten Fink, Dev Gangjee, Brigitte Granville, Dietmar Harhoff, Michal Kazimierczak, Robert Kimble, Georg Licht, David Muls, Pietro Panzarasa, Nathan Wajzman, Kenneth Younge and participants at the USPTO Office of Chief Economist seminar series for useful comments on earlier versions of this paper. Graham acknowledges funding from the Center for International Business Education and Research at Georgia Tech.

<sup>†</sup>Corresponding author: g.v.graevenitz@qmul.ac.uk

<sup>‡</sup>The views expressed are those of the individual authors and do not necessarily reflect official positions of the Office of Chief Economist or the U.S. Patent and Trademark Office.

# 1 Introduction

The scholarly consensus holds that geographical distance affects not only the intensity of trade and migration patterns, but also the diffusion of ideas and innovations (Clark et al., 2018). Kolko (2000) noted that the reduction in communication costs and improvements in the speed and quality of interactions might lead to the “death of distance,” and possibly an end to agglomeration effects, but found evidence only to support the former. Given the intangible nature of ideas, distance may have become less of an impediment to knowledge diffusion after the Internet dramatically lowered communication and interaction costs (Keller and Yeaple, 2013; Head et al., 2018).

Following the early recognition that distance impedes knowledge transfer among people and firms (Marshall, 1920), researchers employed patent data to show that the diffusion of codified knowledge lessens with spatial distance (Henderson et al., 1993).<sup>1</sup> Other researchers have confirmed this distance-impeding effect by exploring the transfer of ideas codified in both patents and scientific publications (Peri, 2005; Belenzon and Schankerman, 2013; Singh and Marx, 2013; Li, 2014). Comin et al. (2012) find that distance impeded diffusion in 20 major technologies over 140 years, while Keller and Yeaple (2013)’s study of trade among corporate affiliates shows that distance lessens knowledge transfer even within firms.

In contrast, Head et al. (2018) rely on more recent data to show that distance is no longer a significant impediment to idea diffusion in the field of mathematics. Using rich data on personal networks of mathematicians, they show that ideas flow primarily through person dyads (i.e., mentor-mentee), and that over time spatial distance, ceterus paribus, has ceased to negatively affect knowledge transfer across such dyads. This finding supports the notion that personal ties are essential for the flow of ideas and innovations, as has also been shown by Singh (2005) and Breschi and Lissoni (2009).

Does this recent “death of distance” in mathematics reflect a more systematic phenomenon across other disciplines and technologies? We investigate the spatial diffusion of innovation across a broad range of industries and firms by exploiting data from the trademark register of the U. S. Patent and Trademark Office (USPTO) to answer this question. This data source allows us to study the diffusion of innovations through the lens of words that identify significant new product and service innovations, for example GPS (global positioning systems), MP3, and the DVD. Specifically, we examine the introduction of new words (tokens) among the 4.5 million words contained in the USPTO trademark register used to describe goods and services during 1980-2012,<sup>2</sup> and identify as “significant innovations” those tokens that are then re-used frequently. We observe both the first occurrence of a token in this word corpus by an “innovator” and its diffusion among the trademark filings of all “follower” firms and entities within one year of first use. Address location data on both innovator and follower firms are available in the records, enabling us to aggregate new-token emergence and diffusion to the year and ZIP code

---

<sup>1</sup>Because patents protect only a subset of all technology innovation and are used selectively, this finding may not be general (Hall and Harhoff, 2012).

<sup>2</sup>The goods and services declarations are largely codified by trademark offices. For instance WIPO provides applicants with the *Madrid Goods & Services Manager*, a tool to help them select appropriate terms and descriptions for the statements that delineate what type of product or service is being protected. However, registrants are not bound to the existing set and can introduce novel descriptors which may in time enter into the official list of descriptors proposed by the offices. We exploit this regularity to study when new terms enter into the trademark register, and how they subsequently diffuse.

level.<sup>3</sup> Accordingly, we are able to examine how geographical distance affects the probability and degree of diffusion in market-based innovation in the United States from 1980-2012.

The paper first provides descriptive results to support the claim that new tokens from the USPTO trademark register measure innovations that are "new to the world" in the sense of the Oslo Manual (OECD and Eurostat, 2018). We present map visualisations to show that such tokens have primarily arisen in densely populated metropolitan areas associated with innovation activity in the previous literature (Peri, 2005; Forman et al., 2016)<sup>4</sup> and that they tend to diffuse primarily to these same areas. The visualisations also show that the 1996-2005 period saw a substantial increase in the establishment of intensive, long-range diffusion links connecting innovators situated in New York, Los Angeles and the San Francisco Bay Area. To further validate trademark tokens as indicators of innovation, we compare the diffusion of several important innovations in patents to diffusion of trademark tokens and confirm that trademark tokens capture substantially the same innovation dynamics for these innovations. This suggests that trademark tokens can capture innovation activity in all sectors and not primarily in manufacturing, as is the case with patents.

To test the effect of distance on innovation diffusion, we construct a panel of 11,587 directed ZIP code-to-ZIP code links over 33 years. We analyze both incidence of first diffusion and the intensity of diffusion between ZIP codes. Building on a simple model of diffusion between ZIP codes, we estimate Instrumental Variables Poisson models of the intensity of diffusion between sending and receiving ZIP codes. These models allow us to address the potential endogeneity of diffusion links and their associated absorptive capacity. Our results show that, for new tokens emerging after 1996, spatial distance no longer affects the likelihood that an innovation will diffuse from its origin ZIP code to a follower ZIP code consistent with Head et al. (2018). However, contingent on previous diffusion from the sending to the receiving ZIP code, we find that geographic distance still exerts a strong, negative affect on the intensity (extent) of diffusion between ZIP codes. And, if anything, this effect has grown stronger over time according to our results. We also find that the degree of innovativeness of a sending ZIP code, and the degree to which the receiving ZIP code has absorbed previous innovations, both determine the establishment of a persistent transfer link between these locations, and increase the intensity at which innovations diffuse between them.

This paper provides a first application of a new measure of innovation drawn from trademark data. We thus contribute to a growing literature on innovation measures derived from a range of novel information sources. Alexopoulos (2011) shows that counts of new books focused on emerging technologies can be used to track the rate of innovation, arguing that innovation measures should i) be available for long periods, in at least annual frequency, ii) be objective, iii) contain innovation dates, iv) reflect the economic importance of innovations, and v) capture a broad range of innovation.

The innovation measure introduced here could in principle be extended as far back as Britain's first trademark register in 1876 (Bently, 2008). In light of our research question we restrict our dataset to the registration years 1980-2012. Since trademarks cover a broad range of goods and services, much wider

---

<sup>3</sup>In the U.S., post codes are called "ZIP" codes. Almost all trademark registrants addresses recorded at the USPTO include ZIP codes; where missing, we geocode street addresses to obtain ZIP codes. While some US ZIP codes have been re-assigned over time, we provide robustness checks of our main analytical results using stable U.S. census tracts to validate our key results.

<sup>4</sup>Patenting intensive regions in the U.S. include New York City, the San Francisco Bay Area, and Greater Los Angeles.

for instance than patents, the measure of innovation we propose reaches an economy-wide breadth similar to the publication measure used by [Alexopoulos \(2011\)](#).<sup>5</sup> Using trademark data to capture innovation has several additional benefits: the trademark register reflects both arrival and diffusion of innovations in a way that can be cleanly established. It is possible to identify firms that originate new tokens, thus allowing us to observe the spatial concentration and distribution of new token arrivals, and to capture temporal and spatial diffusion to "followers" who subsequently use the same token. Moreover, because trademark registration is associated with products and services being introduced in the marketplace, we are arguably focusing on more economically or commercially important tokens than the wider publishing source used in [Alexopoulos \(2011\)](#). Since we focus on new tokens that have been used relatively frequently, we increase the likelihood that we are examining tokens linked to important innovations that have diffused widely, and can also explore how distance affects such diffusion. That said, an innovation measure based on trademark tokens has a key limitation: it may be conservative in that not all innovations may use new tokens in their trademark descriptions. Establishing the significance of this limitation and its temporal evolution will be an important step in determining whether our proposed measure captures the overall amount of innovation adequately.

Since [Schmoch \(2003\)](#) suggested service marks could be used as service innovation indicators, economists have increasingly employed trademark data to provide additional insights into innovation ([Mendonca et al., 2004](#); [Flikkema et al., 2014](#); [Thoma, 2015](#); [Graham et al., 2018](#)). Trademarks are frequently registered close in time to a new product or service being introduced to the market, so are much nearer to launch date than are patents,<sup>6</sup> and more comparable to the publication dates of technical manuals and books. The trademark registers also reflect activity from a much wider set of sectors and firms than U.S. and European patents ([Graham et al., 2013](#); [Dinlersoz et al., 2018](#)). This wider coverage stems from the lower cost of filing trademarks and from their primary objective: to protect a brand or logo against imitation, and to protect consumers from fraud. As such, trademarks are used at least as widely in service industries as they are to sell products. Moreover, because companies selling physical products frequently do not patent ([Moser, 2012](#); [Fontana et al., 2013](#)), trademark data can capture innovation missed in the patent data. Trademarks are also filed nearer in time to market introduction of innovations than patents.

This paper is structured as follows: In the next section we discuss existing measures of innovation and diffusion, introduce our own measure and provide validation. Section 3 contains a discussion of the empirical model we implement and a description of our trademark word-token data set. In Section 4 we present results from estimating the model. Section 5 concludes.

## 2 Measuring Innovation and Diffusion

This section provides a description of innovation and diffusion patterns in the United States after 1980. The section begins by reviewing the literature on the diffusion of ideas and innovations, including prior research on measuring innovation. We then introduce our new trademark measure and describe several

---

<sup>5</sup>One limitation of our indicator re: [Alexopoulos \(2011\)](#) is its basis in trademarks; books were published much earlier than trademark registers were established, providing for a longer time horizon.

<sup>6</sup>In pharmaceuticals, patent filing often precedes product introduction by 7-10 years ([Grabowski and Vernon, 2000](#)).

innovation and diffusion patterns reflected in these data, also providing comparisons of these patterns with those observed in patent data.

## 2.1 Diffusion of Ideas and Innovations

Because patent data has been widely available for decades, analysis of how technological inventions diffuse has commonly relied on patent data, and specifically on patent citation patterns. [Jaffe \(1986\)](#) showed that patent citations could be used to capture knowledge spillovers and [Trajtenberg \(1990\)](#) demonstrated that the number of times a patent is cited captures the technological significance of the patented invention. [Henderson et al. \(1993\)](#) then showed that spillovers of innovations decline with distance.<sup>7</sup> Scientific publications (e.g., journal articles) constitute a separate source of data for studying knowledge diffusion. Publications, like patents, contain citations that may reflect reliance on prior ideas.<sup>8</sup> [Head et al. \(2018\)](#) use rich data from mathematics to demonstrate that distance has ceased to affect the spread of ideas in mathematics. Their finding, based on more recent data, contrasts with the results from most papers on diffusion based on patents which generally employ older data.

Measuring knowledge diffusion using information on technology usage is onerous because links between the innovation and its (ultimate) use must be inferred. Examples include [Comin et al. \(2008\)](#) and [Comin and Hobijn \(2010\)](#), who provide evidence on the diffusion of a wide range of technologies from data constructed to study the adoption of 115 technologies over 200 years, across many countries. [Comin et al. \(2012\)](#) also employ these data and find that distance negatively affects technology diffusion.

While the literature studying the diffusion of ideas and innovations has continued to grow, there are a number of data-related reasons to be careful when interpreting these results. [Nelson et al. \(2014\)](#) argue that firms may over- or underreport adoption of innovations. Moreover, the common use of patent data in these studies can introduce bias since not all sectors or firms rely on patents to protect their innovations ([Levin et al., 1987](#); [Cohen et al., 2000](#)). It is also well established that many patents are not associated with product introductions ([Hall and Harhoff, 2012](#)), which may lead to an overestimate of innovation diffusion, though not possibly of ideas, through patent-based indicators.

In this respect the trademark data we use for our study present some advantages. The USPTO requires that the scope of a mark is restricted to its use in the market. While this requirement is not always met,<sup>9</sup> it increases the strength of the correlation, relative to patenting, between use of a token in a good and service declaration and the actual introduction of a corresponding product in the market. Another benefit of using trademark data to study innovation derives from the low cost of applying for trademarks, thereby increasing the likely range of firms employing and innovations reflected in the trademark register.

A limitation of innovation measures noted by [Nelson et al. \(2014\)](#) may affect our data, however: When new goods and services are introduced there may be considerable variation in the terminology

---

<sup>7</sup>This result has been replicated in a number of recent papers ([Peri, 2005](#); [Belenzon and Schankerman, 2013](#); [Singh and Marx, 2013](#); [Li, 2014](#)).

<sup>8</sup>However, recent research investigates the incidence of so-called negative citations in science ([Catalini et al., 2015](#)).

<sup>9</sup>USPTO conducted a Proof of Use pilot in 2012, requiring trademark owners to submit additional evidence of trademark use on the goods or in connection with the services identified in the registration. In just over half of the randomly selected registrations, owners were unable to verify previously claims of use, resulting in either narrowing of protection through deletion of goods and/or services or outright registration cancellation. See the [Post Registration Proof of Use Pilot Final Report](#) (accessed 12 August 2019).

used to refer to them across firms. It can take time for a commonly accepted term to emerge. This pattern means that we may not correctly identify the date on which some innovations first emerge; rather we may observe the date at which the most widely adopted name for the innovation emerges.<sup>10</sup>

## 2.2 Measures of Innovation

Innovation can be measured for different purposes and at different levels of aggregation, from the firm or unit, up to the economy or world. The Oslo Manual (OECD and Eurostat, 2018), which sets out a standard for collecting and using innovation data, defines an innovation as a new or improved product or service. The most widely used measures of innovation derive from R&D investment data, but these input measures are arguably the least precise measures of innovation. Existing innovation output measures derive from a limited range of sources: Data from the patent registers are widely used to study technological inventions and inventor locations while data from company surveys are widely used to track innovation processes within and across firms and universities. Since patent data reflect innovation in a limited range of technologies, surveys present a more comprehensive tool in their coverage of technologies and industries, but are expensive to conduct and usually reflect activity in only a small subset of firms. Surveys frequently focus on the degree of novelty of innovations, reflecting the approach taken by the Oslo Manual, while studies using patent data largely rely on the idea of the inventive step inherent to the patenting process and typically present patent counts. That said, the degree of novelty inherent in a specific invention is much harder to identify in the patent data.

Recently, new sources of data have produced a richer understanding of innovation. Alexopoulos (2011) and Alexopoulos and Cohen (2011, 2019) exploit data on new book titles covering computers and technology. Hippel et al. (2010) survey consumers in the UK and show that a significant proportion engage in developing and modifying of consumer products. Moser (2012) uses historical data to study innovation beginning in the 1800s when neither patents nor trademarks were widely available. These papers have primarily addressed the question of how much innovation there is (was) as well as when and how innovation arises.

In terms of innovation in recent decades, Forman et al. (2016) study where innovations have arisen in the United States using patent data. They identify an increase in the share of patents originating from the San Francisco Bay area in California, mainly at the expense of the New York City metro area. Much of the shift they identify occurred between 1990 and 2000, yet was not concentrated in ICT technologies: the re-concentration is more general across technologies at least those reflected in patenting. In contrast to a general innovation re-configuration, Hannigan et al. (2015) illustrates the local persistence of automotive innovation in Detroit, Michigan, even as the manufacturing of automobiles has declined in the region. Their analysis documents the simultaneous importance of local and long-distance links to the continued vibrancy of this geographic technology cluster. This paper also relies on patent data as automotive innovations are frequently patented, although survey data suggests the patenting share in automobile technologies may be as low as 49% for product inventions, and 20% for process inventions (Cohen et al., 2000).

The trademark register derived measures of innovation we introduce in this paper cover a much

---

<sup>10</sup>A more focused study of only a small subset of innovations in our data could address this issue, but our approach here is to include as wide a range of innovations as possible.

broader range of industries than the patent data. For instance, trademark protection extends to service industries,<sup>11</sup> an economically vibrant area that patent data largely misses. Moreover, the administrative character of the trademark data means that we are able to link registrant addresses to geographical locations, thus allowing us to study where innovation occurs, and the possible effects of distance on diffusion. The next section provides a first descriptive look at this aspect of our data.

### 2.3 A New Measure of Innovation and Diffusion

We propose to measure innovation through words (tokens) that are entirely new to the corpus of words used to describe products or services in the US trademark register. These we call new to the world tokens at their time of first use. We then extract the subset of tokens that is most frequently re-used within the subsequent five years. Our cutoff for most frequent re-use is those words in the ninth decile of the distribution of re-use frequencies within each Nice class during 1980-2012. We then identify and geocode the addresses of all firms introducing these tokens into the trademark register (hereafter innovators).<sup>12</sup> Figure 1 shows the distribution of these innovators across the United States, subdivided into three periods to show timing effects on the spatial distribution of innovations. Our mapping of trademark-based innovations in Figure 1 comports with findings based on patenting reported in [Forman et al. \(2016\)](#) and [Hannigan et al. \(2015\)](#): the cluster of innovation in and around the San Francisco Bay Area is clearly visible, as is the significant cluster in the Northeast megalopolis from Boston to Washington DC and persistent innovation around the rust-belt cities of Minneapolis, MN, Detroit, MI, Cleveland, OH, Pittsburgh, PA, and Buffalo and Rochester, NY.

We extend our analysis by exploiting subsequent uses of fast growing new tokens in the trademark register, and map the diffusion of these tokens within one year of filing. Figure 2 provides two separate ways of analyzing this diffusion. The first visualizes the locations (left panel) of firms that subsequently used these tokens in another trademark filing (hereafter followers). Followers tend to be proximately located to their matched token-innovator firms, and file very quickly following the first trademark. In fact, a large share of followers are located less than 100 miles from the innovator, and in 15% cases file their trademark within 3 days of the first mark. This finding may provide further evidence of Merton's multiples ([Merton, 1961](#)).

The right panel of Figure 2 focuses on diffusion links that cover distances of at least 100 miles and across which at least 10 innovations have diffused. This graph illustrates the importance of New York, the San Francisco Bay Area and Greater Los Angeles to the supply of innovation in the United States. In this graph of intense innovation diffusion links, the original long-distance diffusion arose primarily between Southern California and New York.

While these figures are interesting and contain far more information than we are able to analyze in this article, it remains to determine whether our measure of innovation is closely aligned with more conventionally used measures of innovation, such as those derived from patent data. The next section addresses this question.

---

<sup>11</sup>Due to a change in the trademark classification system introduced in 2000 we have excluded some of the service classes (42-45) from our analysis.

<sup>12</sup>Geocoding of addresses was done with the help of two Stata packages: `opencagegeo` and `geocodehere`. We geocoded a subset of addresses twice and checked the reliability of the packages used.

# Tokens New to the World

excluding classes beyond 41

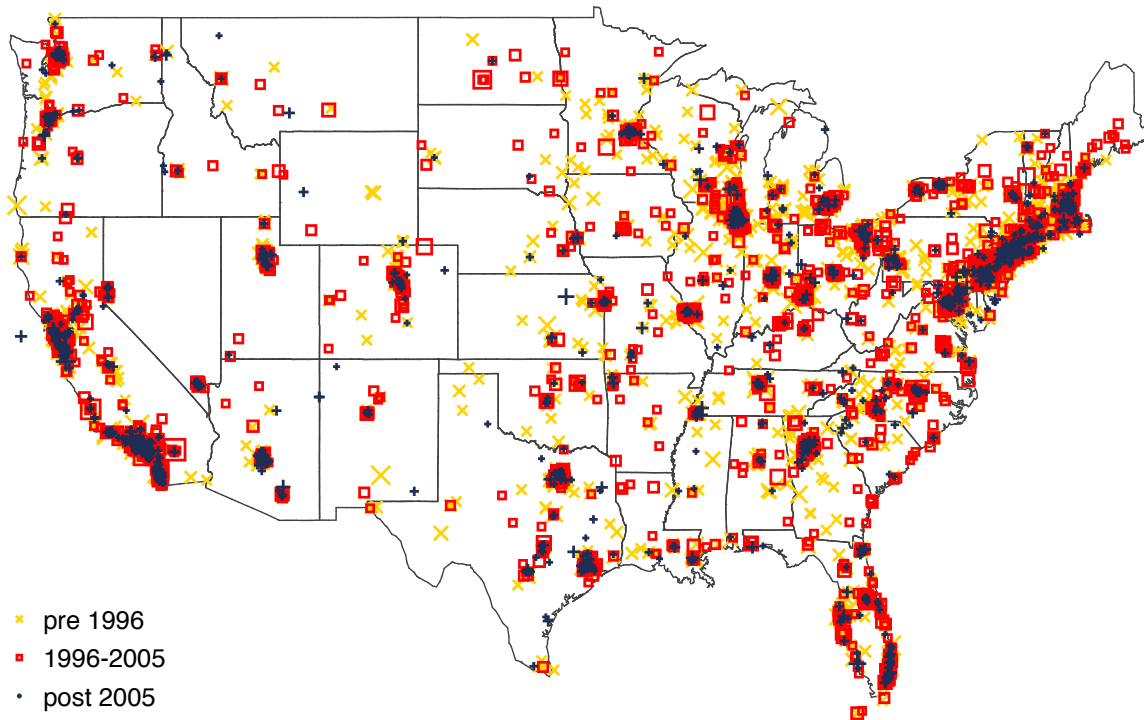


Figure 1

This figure provides a visualization of distinct addresses contributing new to the world word tokens in USPTO trademark descriptions. Symbol size indicates the number of new tokens contributed (larger indicates more tokens) in three distinct time periods (indicated by style and color of symbols). The figure identifies concentrations of innovations in spatial proximity, similar to maps of local labour markets using commuter flows developed by [Nelson and Rae \(2016\)](#). These clusters occur in ten megaregions (e.g. Northeast, Northern California, Southern California) first identified by [Lang and Nelson \(2007\)](#). Due to classification changes, the figure excludes NICE classes 42-45 for consistency.

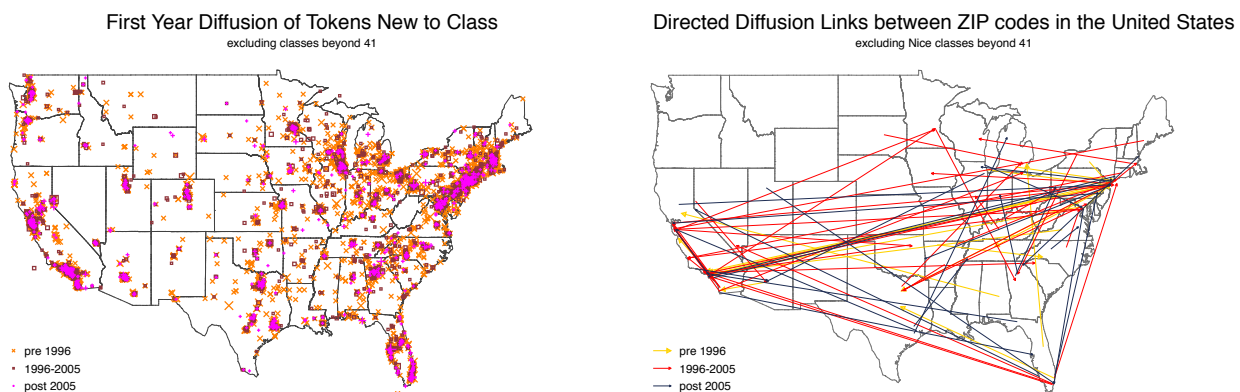


Figure 2: Diffusion of innovations across the USA

This figure shows the diffusion of tokens new to the world within the first year of their introduction in the left panel. The right panel shows arrows starting from addresses at which tokens new to the world are introduced and ending at addresses which subsequently use these tokens within the first year of their introduction. We exclude all distances below 100 miles and only show links across which at least 10 tokens diffused in a given period.



## 2.4 Validation

To validate the use of trademark tokens as measures of innovation and diffusion, we compare the diffusion of four important tokens in the trademark data to diffusion of those same tokens in the titles and abstracts of USPTO patent documents. Figure 3 compares the diffusion patterns for the tokens GPS, Internet, DVD and MP3 that emerged in U.S. patent documents (left panel) with the diffusion patterns of these same tokens found in U.S. trademark registrations (right panel).

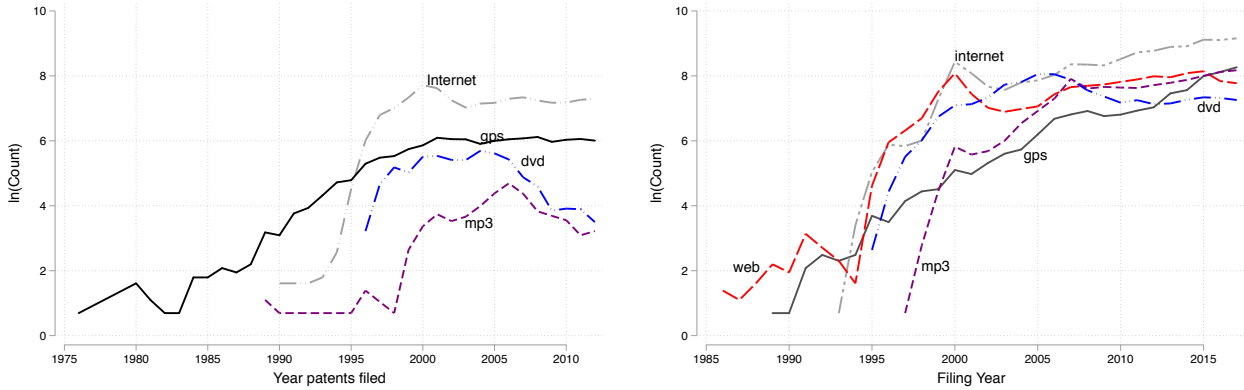


Figure 3: Diffusion of four significant innovations measured using patent tokens (left panel) and trademark tokens (right panel)

The left panel presents counts of patents that contained the tokens in either the title or abstract. Note that the left panel presents data for the years 1975-2012 while the right panel presents data for the years 1985-2015. We extended the left panel back to 1975 in order to represent fully the variation in the GPS series.

A number of qualitative parallels emerge from these two graphs that are worth noting: Use of the tokens Internet and MP3 peaked in 2000 during the internet bubble in both sources. Use of MP3 continued to grow strongly, after a recession induced dip, until around 2006. Turning to Global Positioning System (GPS), it is evident that patenting in GPS technology started many years prior to commercialization, likely due to restrictions imposed by the military. It is also worth pointing out that the diffusion path for GPS in both sources is clearly different from that of the other three technologies, but still similar across the panels. Finally, the decline of patenting related to the DVD technology after 2005 is reflected in a fall in the number of new uses of the DVD tokens for trademarks also. Overall, this descriptive analysis provides us additional confidence that trademark tokens reflect innovation and diffusion in a similar manner to using patent data.

## 3 Model and Data

Henderson et al. (1993, 2005) discuss the primary identification problem that affects all studies seeking to estimate how distance affects diffusion of innovation: Is diffusion localized because distance makes it harder to learn about an innovation, or because those most likely to re-use an innovation are located in the same area as the innovator? In the second case unobserved factors can generate a cluster that is simply revealed by the diffusion patterns observed in the data. Henderson et al. (1993) address the identification

problem by matching citations with potential citations that are comparable.<sup>13</sup> [Head et al. \(2018\)](#) who examine personal ties in mathematics, adopt a similar approach to identification as [Henderson et al. \(1993\)](#) to show that the effects of distance on knowledge diffusion in mathematics have decreased over time.

Due to the novelty of our data source we are not yet in a position to duplicate the micro-level approach to estimating the influence of distance on innovation diffusion. Specifically, we do not have sufficiently detailed information on trademark filing entities and their histories to estimate models at the firm level. Instead, we pursue another approach to the estimation of distance effects on knowledge flows by aggregating our trademark innovation index to the ZIP code level. We then employ these location-level data to estimate gravity models of innovation diffusion between sending and receiving ZIP codes. Previous work by [Peri \(2005\)](#) and [Li \(2014\)](#) used gravity models to show strong negative effects of greater distance on knowledge flows.

After aggregating to localities, we estimate the capacity of locations to generate and absorb innovations, and borrow from the literature modeling international trade flows to analyze a fixed set of regional links. We find that the number of potential sending-and-receiving pairs is far greater than the actual pairs where we observe innovations diffusing.

However, since our data span many years and ZIP codes are generally fixed, we are able to observe repeated diffusion of innovations between locations over time. Rather than relying on matching citing and non-citing ZIP codes to construct controls, we include only those ZIP code pairs where we observe at least one diffusion event but include these pairs for the entire 33 years of our sample. We adopt a gravity-model estimation approach suggested by [Silva and Tenreyro \(2006\)](#) allowing us to keep observations where diffusion counts are zero. We augment this approach by allowing for endogeneity of first diffusion from one location to another by employing lagged variables to instrument the formation of diffusion links between ZIP codes and the absorptive capacity of firms located in the receiving ZIP code.<sup>14</sup>

A further motivation for estimating gravity-type models at the ZIP code level can be found in the literature on regional innovation systems and clusters. [Hannigan et al. \(2015\)](#) suggest that capabilities which spawn innovation reside in local and global linkages between clusters. These capabilities endure, and can survive the death or migration of specific entities like firms or research centers which may help to explain our showing of persistent innovation clusters in older rust-belt cities like Detroit and Buffalo.

### 3.1 Model

We adopt a model of innovation diffusion comparable in spirit to the model proposed by [Peri \(2005\)](#). We posit that diffusion  $d_{s,r,t}$  from a sending ( $s$ ) to a receiving ( $r$ ) area in year  $t$  is a function of distance between these areas  $D_{s,r}$ , innovation  $I_{s,t}$  in the sending location and absorptive capacity  $A_{r,t}$  in the receiving location and area, time and technology fixed effects  $X_{s,r,t,n}$ :

$$d_{s,r,t} = X_{s,r,t,n} (D_{s,r})^{\beta_D} (I_{s,t})^{\beta_I} (A_{r,t})^{\beta_A} . \quad (1)$$

<sup>13</sup> This approach was subsequently critically tested by [Thompson and Fox-Kean \(2005\)](#) and commented on in [Henderson et al. \(2005\)](#). [Singh and Marx \(2013\)](#) extend the methodology used to estimate this type of matching model.

<sup>14</sup> Methods used to endogenise variables in a Poisson framework are developed by [Windmeijer and Santos Silva \(1997\)](#).

This specification can be estimated using a Poisson model. In order to address the identification problem with which the micro-level literature has grappled (Henderson et al., 1993), we endogenise the first instance of diffusion from sending to receiving location and control for the age of the link between them. We also allow for the endogeneity introduced by a receiving areas absorptive capacity, but assume that the arrival of new to the world innovations at the sending location is uncorrelated with ZIP code-pair specific unobserved effects.<sup>15</sup>

This leads us to estimate the following Poisson model with endogenous covariates and an additive error term (Windmeijer and Santos Silva, 1997; Silva and Tenreyro, 2006):

$$\delta_{s,r,t} = \exp(\beta_0 + \beta_D \ln D_{s,r,t} + \beta_I \ln I_{s,t} + \beta_A \ln A_{r,t} + \gamma_L L_{s,r,t} + F'_{s,r,t,n} \lambda) + u_{s,r,t}, \quad (2)$$

where  $\delta_{s,r,t}$  is a count of the number of tokens introduced as new to the world in year  $t$  in the sending location which are used within one year at the receiving location. This forms our index of innovation diffusion  $d_{s,r,t}$ .

$\ln D_{s,r,t}$  is the logarithm of the median distance between sending and receiving addresses for each pair of ZIP codes and can vary across years within each pair,  $\ln I_{s,t}$  is the logarithm of the innovation count in the sending region and  $\ln A_{r,t}$  is the count of all terms diffusing to the receiving area net of those from the sending area.  $L_{s,r,t} = 1$  identifies the first year of diffusion from the sending to the receiving location. When estimating this model we control for sending and receiving ZIP code fixed effects at the two digit level as well as trademark Nice-class fixed effects and year fixed effects, all subsumed under  $F_{s,r,t,n}$ .

We model the probability of first diffusion ( $L_{s,r,t}$ ) using lagged innovative potential and lagged absorptive capacity of receiving areas and lagged absorptive capacity of sending areas. The evolution of current absorptive capacity of the receiving area ( $A_{r,t}$ ) is modeled using the same instruments.

## 3.2 Data

Using the USPTO trademark data we are able to analyze the appearance of new tokens after 1979. Because we are interested in important innovations we define significant new tokens as those in the top diffusion decile (by re-use frequency, within Nice class in a five-year forward window) of the distribution of all new tokens, and identify 28,686 significant new tokens in our data. These significant tokens originate from addresses in 3,523 distinct ZIP codes and are received by (diffuse to) addresses in 5,370 distinct ZIP codes. In 17% of these cases, the sending and receiving addresses lie within the same ZIP code.<sup>16</sup>

We construct local area aggregates of new to the world token introductions and their diffusion using ZIP code locations. This method allows us to generate location aggregates by year.<sup>17</sup> We identify 11,587

<sup>15</sup>The latter assumption may be quite strong. We cannot test this as estimation of Instrumental Variables Poisson models is challenging with large numbers of endogenous variables. In future work we may be able to relax this assumption, once a larger range of years in which diffusion takes place can be accommodated. For instance, extending the diffusion period from one year (as is now) to five years requires geocoding of twelve times more addresses than our present data contains.

<sup>16</sup>This does not affect our analysis of distance effects as we construct the distance between sending and receiving addresses per ZIP code pair as the median distance within each pair and year. As we have geocoded the addresses of all filing entities we are able to construct distances even within the same ZIP code.

<sup>17</sup>ZIP codes can change over time. This has the effect of splitting records of ZIP code pairs into two or more separate

pairs of ZIP codes having at least one instance of new to the world token diffusion between a sending and a receiving ZIP code, and construct a panel for all these pairs covering 33 years. Table 1 sets out several descriptive statistics.

Table 1: **Descriptive Statistics** (N = 382,364)

Variable	mean	sd	median	min	max
Diffusion <sub>s,r,t</sub>	0.052	0.606	0	0	158
ln Net Diffusion <sub>r,t</sub>	0.011	0.139	0	0	5.242
Link dummy	0.030	—	0	0	1
ln Distance <sub>s,r,t</sub>	5.659	2.560	6.725	0	9.335
ln Innovation <sub>s,t</sub>	0.041	0.196	0	0	4.443
ln Years linked <sub>s,r,t</sub>	0.891	1.153	0	0	3.497
ln Net Diffusion <sub>s,t-1</sub>	0.018	0.195	0	0	5.602
ln Net Diffusion <sub>r,t-1</sub>	0.006	0.105	0	0	5.242
ln Innovation <sub>r,t-1</sub>	0.077	0.258	0	0	4.443

The dependent variable of our main model is the count of diffusion events between sending and receiving ZIP codes per year: Diffusion<sub>s,r,t</sub>. Our model includes an endogenous dummy variable that is set to one when significant innovations diffuse from a sending to a receiving ZIP code for the first time. The models we estimate contain three principal variables:

Distance<sub>s,r,t</sub> in miles calculated as the median distance of all sending - receiving firm pairs per year for each ZIP code pair. Distance can vary within a pair of ZIP codes over different years to reflect changes in the concentration of economic activity in different locations within each ZIP code. Each individual distance is the geodesic distance between the sending and receiving locations, calculated using the haversine formula. The literature suggests that this variable will have a negative effect on diffusion.

Innovation<sub>s,t</sub> is the count of all new to the world tokens generated in the sending ZIP code per year. We also use the lagged count of innovation per year in the receiving ZIP code as an instrument. Innovation is treated as being exogenous throughout.

Net diffusion<sub>r,t</sub> is the count of all new to the world tokens diffusing to a receiving ZIP code, net of those from a sending ZIP code per year. This variable can be thought of as a proxy for the absorptive capacity of the firms within the receiving ZIP code location (Cohen and Levinthal, 1990). We instrument this variable in our main specifications.

---

records in our data. We have identified cases in which there are multiple ZIP code pairs for the same locations and test whether these cases affect our conclusions. We find no significant differences. The results of this robustness check are relegated to the Appendix, Table 5.

In addition, we include several covariates:

Years linked $_{s,r,t}$  measures the number of years that have passed since the first diffusion of an innovation from the sending to the receiving ZIP code.

Period has three phases: before 1996, between 1996 and 2005 and after 2005. The phases are chosen so as to separate out the decade centered on the Dot com boom in 2000, during which significant investments in Internet mediated communication took place and U.S. trademark registrations were unusually high (Graham et al., 2013).

## 4 Results

This section contains estimation results obtained from estimating the empirical model set out in Section 3.1 with the data described in the previous section. We begin with a discussion of the two first stage models and then present and discuss the main results. Appendix, Section B contains further results, which are obtained by estimating the main model without instrumenting the link dummy and net diffusion to the receiving ZIP codes.

### 4.1 First Stage Models: Probability of Diffusion and Net Diffusion

Table 2 provides two sets of results. The first is obtained from estimation of a model for the probability that an innovation new to the world diffuses from a sending ZIP code to a specific receiving ZIP code in a specific year. This model endogenizes the instance of first diffusion between two ZIP codes. The second set of results is obtained from estimating a model for the net rate of diffusion to a receiving ZIP code in a specific year.

These models all contain the distance between sending and receiving ZIP codes as well as the innovation rate in the sending ZIP code. We estimate two versions of each model: The first contains all variables listed in the table, while the second allows for interactions between distance and the time period. We introduce three variables to instrument the endogenous variables: lagged net diffusion to the receiving ZIP code, lagged net diffusion to the sending ZIP code, and lagged innovation in the receiving ZIP code.

The probability that a new to the world innovation arising in a sending ZIP code diffuses to a receiving ZIP code decreases with the median distance between all sending and receiving firms in the two locales in a given year. While this effect is statistically significant, it is small in absolute value: An increase in the distance from zero to the maximum distance recorded in the data between any two locations reduces the probability of diffusion by only 0.2% (Model 1). Moreover, the interaction with dummies for periods two (1996-2005) and three (post 2005) results in insignificant effects for distance in those two periods (Model 2). These results support the notion that distance is not, or is no-longer, an meaningful impediment to innovation diffusion between even fairly remote areas in the United States. This finding is notable given the countrys large land area (9.8 million km<sup>2</sup>).

Table 2 also shows that innovation in either the sending or receiving ZIP code increases the probability that innovations will diffuse between them. Interestingly however, a higher rate of diffusion in the past from other ZIP codes to the receiving ZIP code reduces that probability of diffusion, suggesting

that persistent transfer links between specific locales may pose barriers to entry to ideas and innovations from different locations.

Table 2: First Stage Models: First Diffusion and Net Diffusion

Dependent Variable Distance & Periods	First Diffusion		Net Diffusion	
	in levels	interacted	in levels	interacted
	(1)	(2)	(3)	(4)
$\ln \text{Distance}_{s,r,t}$	-0.0002*** (0.0000)	-0.0006*** (0.0001)	-0.0003*** (0.0001)	-0.0005*** (0.0001)
$\ln \text{Innovation}_{s,t}$	0.2509*** (0.0092)	0.2509*** (0.0092)	0.0931*** (0.0050)	0.0931*** (0.0050)
$\ln \text{Net Diffusion}_{r,t-1}$	-0.0065* (0.0026)	-0.0065* (0.0026)	0.0071* (0.0030)	0.0071* (0.0030)
$\ln \text{Innovation}_{r,t-1}$	0.0133*** (0.0020)	0.0133*** (0.0020)	0.0091*** (0.0012)	0.0092*** (0.0012)
$\ln \text{Net Diffusion}_{s,t-1}$	-0.0017 (0.0022)	-0.0017 (0.0022)	0.0003 (0.0015)	0.0002 (0.0015)
Period(= 2) $\times \ln \text{Distance}_{s,r,t}$		0.0008** (0.0003)		0.0000 (0.0002)
Period(= 3) $\times \ln \text{Distance}_{s,r,t}$		0.0009** (0.0003)		0.0007*** (0.0002)
Constant	0.0145*** (0.0030)	0.0169*** (0.0030)	-0.0034 (0.0021)	-0.0025 (0.0021)
Observations	382364	382364	382364	382364
F-Test	141.07	140.11	50.95	50.66

<sup>1</sup> Robust standard errors clustered at ZIP code level 2 in parentheses: +  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

<sup>2</sup> All models include fixed effects for Nice class, year, sending and receiving ZIP code at level 2.

The first stage models for net diffusion to the receiving ZIP code is largely similar to the models just discussed. There is one clear difference: Net diffusion to a receiving locale is a positive function of its own lag. This positive feedback outcome is to be expected.

## 4.2 Instrumental Variables Models for Diffusion of Innovations

Table 3 contains four models for the level of diffusion between a sending and a receiving ZIP code. The first two models presented do not control for the time passed between a first instance of diffusion from sending to receiving ZIP code. Models 2 and 4 allow for interactions between distance and time periods. The table also provides marginal effects for the distance variable, in each time period, below the main results. Table 4 contains analogous results from models in which we neither instrument the

first diffusion dummy nor net diffusion to the receiving ZIP code.

Table 3: IV Poisson Models for Diffusion of Innovation

Distance & Periods	in levels	interacted	in levels	interacted
	(1)	(2)	(3)	(4)
First Diffusion (1/0)	1.3321 (1.4218)	1.3095 (1.3786)	-0.0750 (1.7232)	-0.0692 (1.7325)
ln Net Diffusion <sub>r,t</sub>	1.9293** (0.6962)	2.0072** (0.6547)	2.3561*** (0.6414)	2.4231*** (0.6614)
ln Distance <sub>s,r,t</sub>	-0.1405*** (0.0254)	-0.1598*** (0.0191)	-0.1496*** (0.0297)	-0.1521*** (0.0295)
ln Innovation <sub>s,t</sub>	1.4470** (0.4972)	1.4437** (0.4974)	1.4418** (0.4421)	1.4379** (0.4520)
ln Years linked <sub>s,r,t</sub>			-2.2294*** (0.4573)	-2.2137*** (0.4852)
Constant	-4.1320*** (0.5482)	-4.0576*** (0.4946)	-3.6898*** (0.5984)	-3.6339*** (0.5834)
Observations	382364	382364	382364	382364
Marginal Effects				
ln Distance <sub>pre 1996</sub>	-0.0047*** (0.0009)	-0.0053*** (0.0007)	-0.0049*** (0.0011)	-0.0050*** (0.0011)
ln Distance <sub>1996-2005</sub>	-0.0114*** (0.0020)	-0.0091** (0.0028)	-0.0123*** (0.0025)	-0.0100** (0.0032)
ln Distance <sub>post 2005</sub>	-0.0077*** (0.0016)	-0.0106*** (0.0028)	-0.0081*** (0.0020)	-0.0128** (0.0045)

<sup>1</sup> Robust standard errors clustered at ZIP code level 2 in parentheses: <sup>+</sup>  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

<sup>2</sup> All models include fixed effects for Nice class, year, sending and receiving ZIP code at level 2.

<sup>3</sup> Instruments: ln Innovation<sub>r,t-1</sub>, ln Net Diffusion<sub>r,t-1</sub>, ln Net Diffusion<sub>s,t-1</sub>

In contrast to the first stage models, these results show that distance has a statistically significant large and negative effect on the number of innovations that diffuse from a sending to a receiving ZIP code: A one-standard deviation change in the logarithm of distance reduces the count of diffusion events by approximately 50% relative to the mean during the post-2005 period. Also notable is that instrumenting first diffusion and net diffusion to the receiving ZIP code increases the estimated coefficient on distance by 12%.

Table 3 shows that net diffusion to a receiving ZIP code increases the number of innovations generated in the sending ZIP code that diffuse to the receiving ZIP code (suggesting an absorptive capacity pull effect) as does an increasing number of innovations arising in the sending ZIP code (suggesting an

innovation supply effect). Coefficients on both of these variables are severely downward biased when we do not instrument first diffusion and net diffusion to the receiving ZIP code. Our results also show that the count of innovations diffusing from sending to receiving ZIP codes tends to fall over time, as the diffusion links between ZIP code pairs age.

To summarize, our results show that spatial distance no longer affects the creation of diffusion links after 1996. However, contingent on previous diffusion from a sending to a receiving ZIP code, we find persistent, strong and negative effects of greater spatial distance on the intensity (extent) of diffusion for existing transfer links between locations.

## 5 Conclusion

This paper contains new evidence on the effect that geographic distance has on the diffusion of innovations. A primary contribution is the description and application of a previously unused source of information on innovations and their diffusion, namely the emergence and re-use of new to the world terms (tokens) contained in the goods and services descriptions of administrative trademark registrations. While this paper considers U.S. trademark information generated during only three decades, there is wide scope for this measure to be constructed from public trademark information in any country, and for periods beginning as early as the late-1800s.

While the consensus scholarly view is that the diffusion of ideas and innovation decreases as spatial distance increases, recent scholarship questions this relationship. By linking new trademark tokens to the business addresses of innovator and follower firms, and defining substantial innovations as new to the world tokens in goods and services classes that are most often transferred between locations, we are able to analyze the diffusion patterns of the most impactful innovations from 1980 through 2012.

Our results largely confirm findings in previous work, which has shown that distance hampers the spread of innovations and ideas. The novelty of our findings lies in the source of information about innovation and diffusion, which is entirely different from that exploited in previous work. We address some of the endogeneity likely to affect this type of work. However, we do not have experimental or quasi-experimental data at our disposal and are not in a position to control for the personal networks which may mediate the diffusion of innovations. Therefore we cannot entirely rule out that diffusion is primarily local, because even in the age of the Internet, the networks of innovators and gatekeepers in the sense of [Roberts and Fusfeld \(1980\)](#) remain local networks.

Trademark tokens, as described in this paper, open opportunities for research into the diffusion of innovations. The trademark token data are particularly useful for analysis of technologies, industries, firms, and economies for which patents - the most common data used to analyze diffusion - are less suitable. Accordingly, our descriptive analysis of the trademark token data suggests that further work is warranted to better understand what additional insights trademark tokens can reveal about diffusion patterns post 1876. If it can be established that the propensity for new tokens describing radical new technologies has remained relatively constant over the period after 1876, trademark tokens would be useful indicators of the amount of innovation generated since. Such a finding would help to answer many questions, including those being raised in the recent literature about the productivity slowdown affecting advanced economies ([Gordon, 2018](#); [Alexopoulos and Cohen, 2019](#)).



## References

- ALEXOPOULOS, M. (2011): “Read All about It!! What Happens Following a Technology Shock?” *The American Economic Review*, 101, 1144–1179.
- ALEXOPOULOS, M. AND J. COHEN (2011): “Volumes of Evidence: Examining Technical Change in the Last Century through a new Lens,” *Canadian Journal of Economics/Revue canadienne d'économique*, 44, 413–450.
- (2019): “Will the New Technologies Turn the Page on US Productivity Growth?” *Economics Letters*, 175, 19–23.
- BELENZON, S. AND M. SCHANKERMAN (2013): “Spreading the Word: Geography, Policy, and Knowledge Spillovers,” *Review of Economics and Statistics*, 95, 884–903.
- BENTLY, L. (2008): “The Making of Modern Trade Mark Law: The Construction of the Legal Concept of Trade Mark (1860–1880),” in *Trade Marks and Brands: An Interdisciplinary Critique*, ed. by L. Bently, J. Davis, and J. C. Ginsburg, Cambridge: Cambridge University Press, chap. 1, 3–41.
- BRESCHI, S. AND F. LISSONI (2009): “Mobility of Skilled Workers and Co-invention Networks: An anatomy of Localized Knowledge Flows,” *Journal of Economic Geography*, 9, 439–468.
- CATALINI, C., N. LACETERA, AND A. OETTL (2015): “The incidence and role of negative citations in science,” *Proceedings of the National Academy of Sciences*, 112, 13823–13826.
- CLARK, G. L., M. P. FELDMAN, M. S. GERTLER, AND D. WÓJCIK (2018): *The New Oxford Handbook of Economic Geography*, Oxford University Press.
- COHEN, W. M. AND D. A. LEVINTHAL (1990): “Absorptive Capacity: A New Perspective on Learning and Innovation,” *Administrative science quarterly*, 35, 128–152.
- COHEN, W. M., R. R. NELSON, AND J. P. WALSH (2000): “Protecting their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (or Not),” Working Paper 7552, NBER.
- COMIN, D. AND B. HOBIJN (2010): “An Exploration of Technology Diffusion,” *American Economic Review*, 100, 2031–59.
- COMIN, D., B. HOBIJN, AND E. ROVITO (2008): “A new approach to measuring technology with an application to the shape of the diffusion curves,” *The Journal of Technology Transfer*, 33, 187–207.
- COMIN, D. A., M. DMITRIEV, AND E. ROSSI-HANSBERG (2012): “The Spatial Diffusion of Technology,” Tech. rep., National Bureau of Economic Research.
- DINLERSOZ, E. M., N. GOLDSCHLAG, A. MYERS, N. ZOLAS, ET AL. (2018): “An Anatomy of US Firms Seeking Trademark Registration,” Tech. rep.

- FLIKKEMA, M., A.-P. DE MAN, AND C. CASTALDI (2014): “Are trademark counts a valid indicator of innovation? Results of an in-depth study of new benelux trademarks filed by SMEs,” *Industry and Innovation*, 21, 310–331.
- FONTANA, R., A. NUVOLARI, H. SHIMIZU, AND A. VEZZULLI (2013): “Reassessing patent propensity: Evidence from a dataset of R&D awards, 1977–2004,” *Research Policy*, 42, 1780–1792.
- FORMAN, C., A. GOLDFARB, AND S. GREENSTEIN (2016): “Agglomeration of Invention in the Bay Area: Not just ICT,” *American Economic Review*, 106, 146–51.
- GORDON, R. J. (2018): “Declining American economic growth despite ongoing innovation,” *Explorations in Economic History*, 69, 1–12.
- GRABOWSKI, H. G. AND J. M. VERNON (2000): “Effective Patent Life in Pharmaceuticals,” *International Journal of Technology Management*, 19, 98–120.
- GRAHAM, S. J. H., G. HANCOCK, A. C. MARCO, AND A. F. MYERS (2013): “The USPTO Trademark Case Files Dataset: Descriptions, Lessons, and Insights,” *Journal of Economics & Management Strategy*, 22, 669–705.
- GRAHAM, S. J. H., A. C. MARCO, AND A. F. MYERS (2018): “Monetizing Marks: Insights from the USPTO Trademark Assignment Dataset,” *Journal of Economics & Management Strategy*, 27, 403–432.
- HALL, B. H. AND D. HARHOFF (2012): “Recent Research on the Economics of Patents,” Working Paper 17773, National Bureau of Economic Research.
- HANNIGAN, T. J., M. CANO-KOLLMANN, AND R. MUDAMBI (2015): “Thriving Innovation Amidst Manufacturing Decline: The Detroit Auto Cluster and the Resilience of Local Knowledge Production,” *Industrial and Corporate Change*, 24, 613–634.
- HEAD, K., Y. A. LI, AND A. MINONDO (2018): “Geography, Ties, and Knowledge Flows: Evidence from Citations in Mathematics,” *Review of Economics and Statistics*.
- HENDERSON, R., A. JAFFE, AND M. TRAJTENBERG (2005): “Patent Citations and the Geography of Knowledge Spillovers: A Reassessment: Comment,” *American Economic Review*, 95, 461–464.
- HENDERSON, R., A. B. JAFFE, AND M. TRAJTENBERG (1993): “Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations,” *Quarterly Journal of Economics*, 108, 577–598.
- HIPPEL, E. V., J. D. JONG, AND S. FLOWERS (2010): “Comparing Business and Household Sector Innovation in Consumer Products: Findings from a Representative Study in the UK,” *Social Science Research Network*, 1–39.
- JAFFE, A. B. (1986): “Technological Opportunity and Spillovers of R&D: Evidence from Firms’ Patents, Profits and Market Value,” *American Economic Review*, 76, 984–1001.

- KELLER, W. AND S. R. YEAPLE (2013): “The Gravity of Knowledge,” *American Economic Review*, 103, 1414–44.
- KOLKO, J. (2000): *The Internet Upheaval: Raising Questions, Seeking Answers in Communications Policy*, MIT press Cambridge, MA, chap. 4.: The Death of Cities? The Death of Distance? Evidence from the Geography of Commercial Internet Usage, 73–98.
- LANG, R. E. AND A. C. NELSON (2007): “America 2040: The Rise of the Megapolitans,” *Planning*, 73, 7–12.
- LEVIN, R. C., A. K. KLEBORICK, R. R. NELSON, AND S. G. WINTER (1987): “Appropriating the Returns from Industrial Research and Development,” *Brookings Papers on Economic Activity*, 3.
- LI, Y. A. (2014): “Borders and Distance in Knowledge Spillovers: Dying over Time or Dying with Age? - Evidence from Patent Citations,” *European Economic Review*, 71, 152–172.
- MANSON, S., J. SCHROEDER, D. VAN RIPER, S. RUGGLES, ET AL. (2019): “IPUMS National Historical Geographic Information System: Version 13.0 [Database],” *Minneapolis: University of Minnesota*.
- MARSHALL, A. (1920): *Principles of Economics*, Macmillan London.
- MENDONCA, S., T. S. PEREIRA, AND M. M. GODINHO (2004): “Trademarks as an Indicator of Innovation and Industrial Change,” *Research Policy*, 33, 1385 – 1404.
- MERTON, R. K. (1961): “Singletons and Multiples in Scientific Discovery: A Chapter in the Sociology of Science,” *Proceedings of the American Philosophical Society*, 105, 470–486.
- MOSER, P. (2012): “Innovation without Patents: Evidence from Worlds Fairs,” *The Journal of Law and Economics*, 55, 43–74.
- NELSON, A., A. EARLE, J. HOWARD-GRENVILLE, J. HAACK, AND D. YOUNG (2014): “Do Innovation Measures Actually Measure Innovation? Obliteration, Symbolic Adoption, and Other Finicky Challenges in Tracking Innovation Diffusion,” *Research Policy*, 43, 927–940.
- NELSON, G. D. AND A. RAE (2016): “An Economic Geography of the United States: From Commutes to Megaregions,” *PloS One*, 11, e0166083.
- OECD AND EUROSTAT (2018): *Oslo Manual 2018*.
- PERI, G. (2005): “Determinants of Knowledge Flows and their Effect on Innovation,” *Review of Economics and Statistics*, 87, 308–322.
- ROBERTS, E. B. AND A. R. FUSFELD (1980): “Critical functions: Needed Roles in the Innovation Process,” .
- SCHMOCH, U. (2003): “Service Marks as Novel Innovation Indicator,” *Research Evaluation*, 12, 149–156.

- SILVA, J. S. AND S. TENREYRO (2006): “The Log of Gravity,” *The Review of Economics and statistics*, 88, 641–658.
- SINGH, J. (2005): “Collaborative Networks as Determinants of Knowledge Diffusion Patterns,” *Management Science*, 51, 756–770.
- SINGH, J. AND M. MARX (2013): “Geographic Constraints on Knowledge Spillovers: Political Borders vs. Spatial Proximity,” *Management Science*, 59, 2056–2078.
- THOMA, G. (2015): “The Value of Patent and Trademark Pairs,” in *Academy of Management Proceedings*, Academy of Management Briarcliff Manor, NY 10510, vol. 2015, 12373.
- THOMPSON, P. AND M. FOX-KEAN (2005): “Patent Citations and the Geography of Knowledge Spillovers: A Reassessment,” *American Economic Review*, 95, 450–460.
- TRAJTENBERG, M. (1990): “A Penny for your Quotes: Patent Citations and the Value of Innovations,” *Rand Journal of Economics*, 21, 172–187.
- WINDMEIJER, F. A. AND J. M. SANTOS SILVA (1997): “Endogeneity in Count Data Models: An Application to Demand for Health Care,” *Journal of Applied Econometrics*, 12, 281–294.

## A Innovation and Diffusion Indicators using Trademark Tokens

All trade mark applications contain a description of the goods and services which the mark will be used for. These descriptions contain widely used terms that are familiar to both the examiners at USPTO and trade mark attorneys. The use of such standard terms simplifies disputes about overlap of goods and services and makes translation easier. We refer to the list of all terms used to describe goods and services as a corpus of tokens and exploit the introduction of new tokens into this corpus<sup>18</sup>.

Goods and services descriptors attached to each trade mark application filed at USPTO provide a good approximation to the range of products marketed under the trade mark name. This is due to the requirement that applicants provide USPTO with proof of the use of their marks in commerce as they are described in the application. When applicants introduce entirely new types of products or services, e.g. GPS, they will often also need to introduce new tokens into the corpus. Should the market for such new products and services grow we would expect many firms entering these markets and filing marks to take up new tokens in their own filings. In order to identify innovation we study the introduction of those new tokens in the corpus that are subsequently widely adopted.

From data about such new, fast growing terms we extract the fastest growing 10% applying an algorithm with four steps:

1. determine the year in which new tokens first appear in the corpus of all tokens in a given Nice class;

---

<sup>18</sup>In natural language processing a collection of texts used as a basis for a descriptive analysis is referred to as a *corpus* of text. The term *token* is used to refer to individual words within such a corpus.

2. obtain the frequency with which new tokens are used in the first five years within that Nice class;
3. obtain the 9th decile of the new tokens frequency distribution, where this distribution is constructed relative to the Nice class and the sub-corpus of all new tokens in that Nice class across all years;
4. retain those tokens used more frequently than the 9th decile in the distribution of frequencies.

This algorithm contains a number of parameters that can be adjusted, such as the number of years over which the impact of the new tokens is measured and the quantile used to define significant tokens. Exploration of variants is left to future work.

We restrict our sample of diffusion events to subsequent uses of each significant token within the first year (365 days not calendar year) of its introduction into the corpus. This restriction is adopted to limit the extent to which addresses had to be geocoded. It also has the benefit of limiting the likelihood that diffusion takes place indirectly.

## B Further Results

Table 4: Poisson Models for Diffusion of Innovation

Distance & Periods	in levels (1)	interacted (2)	in levels (3)	interacted (4)
First Diffusion (1/0)	4.9500*** (0.1870)	4.9486*** (0.1876)	4.3737*** (0.1841)	4.3715*** (0.1850)
ln Net Diffusion <sub>r</sub>	0.0781 (0.0484)	0.0793 (0.0486)	-0.1998 (0.2540)	-0.1972 (0.2542)
ln Distance	-0.1291*** (0.0046)	-0.1327*** (0.0101)	-0.1277*** (0.0046)	-0.1356*** (0.0099)
ln Innovation <sub>s</sub>	0.9701*** (0.0858)	0.9705*** (0.0863)	0.9836*** (0.0949)	0.9850*** (0.0954)
ln Years linked			-1.1776*** (0.1944)	-1.1786*** (0.1943)
Constant	-6.0315*** (0.3256)	-5.9998*** (0.3298)	-5.6613*** (0.3082)	-5.6109*** (0.3088)
Observations	382364	382364	382364	382364

<sup>1</sup> Robust standard errors clustered at ZIP code level 2 in parentheses: <sup>+</sup>  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

<sup>2</sup> All models include fixed effects for Nice class, year, sending and receiving ZIP code at level 2.

This section contains additional results. These support the main results presented in the paper. There are three tables below: Table 4 contains results from estimating Poisson models without instrumenting the

probability of link formation or absorptive capacity at the receiving ZIP code; Table 5 contains results from re-estimating the models presented in Table 3 with two additional dummies capturing cases in which ZIP codes are altered and Table 6 contains results from estimating Instrumental Variables Poisson models for which counts of innovations and diffusion events are aggregated to census tracts rather than ZIP codes. We briefly comment on the results contained in each table here.

Table 5: Robustness Check - IV Poisson Models for Diffusion of Innovation

Distance & Periods	in levels	interacted	in levels	interacted
	(1)	(2)	(3)	(4)
First Diffusion (1/0)	1.3111 (1.3990)	1.3246 (1.3553)	-0.0977 (1.6891)	-0.1130 (1.7021)
ln Net Diffusion <sub>r,t</sub>	1.9220** (0.6766)	1.9815** (0.6470)	2.3446*** (0.6249)	2.4103*** (0.6407)
ln Distance <sub>s,r,t</sub>	-0.1408*** (0.0245)	-0.1603*** (0.0181)	-0.1498*** (0.0290)	-0.1526*** (0.0284)
ln Innovation <sub>s,t</sub>	1.4598** (0.4896)	1.4478** (0.4872)	1.4522*** (0.4324)	1.4538*** (0.4403)
ln Years linked <sub>s,r,t</sub>			-2.2483*** (0.4460)	-2.2388*** (0.4686)
Change ZIP <sub>r</sub>	0.2034 (0.2264)	0.2019 (0.2268)	0.3186 (0.2374)	0.3120 (0.2445)
Change ZIP <sub>s</sub>	0.3498 (0.2695)	0.3796 (0.2676)	0.6047* (0.2465)	0.6455* (0.2597)
Constant	-4.1342*** (0.5386)	-4.0646*** (0.4906)	-3.7035*** (0.5902)	-3.6510*** (0.5708)
Observations	382364	382364	382364	382364
Marginal Effects				
ln Distance <sub>pre 1996</sub>	-0.0047*** (0.0009)	-0.0054*** (0.0007)	-0.0049*** (0.0010)	-0.0050*** (0.0010)
ln Distance <sub>1996-2005</sub>	-0.0115*** (0.0020)	-0.0091*** (0.0027)	-0.0124*** (0.0024)	-0.0100** (0.0031)
ln Distance <sub>post 2005</sub>	-0.0077*** (0.0015)	-0.0105*** (0.0027)	-0.0081*** (0.0020)	-0.0127** (0.0044)

<sup>1</sup> Robust standard errors clustered at ZIP code level 2 in parentheses: +  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

<sup>2</sup> All models include fixed effects for Nice class, year, sending and receiving ZIP code at level 2.

<sup>3</sup> Instruments: ln Innovation<sub>r,t-1</sub>, ln Net Diffusion<sub>r,t-1</sub>, ln Net Diffusion<sub>s,t-1</sub>

Table 4 shows that even in a "naive" Poisson model distance has a statistically significant negative

effect on diffusion of innovations. The coefficients presented in the table are around 12% below those we obtain when instrumenting the probability of first diffusion between two ZIP codes and the absorptive capacity of the receiving ZIP code (cf. Table 3).

Table 4 also shows that our proxy for absorptive capacity ( $\text{Net Diffusion}_r$ ) at the receiving location is not statistically significant in these "naive" models. This is because the coefficients are very significantly downward biased when compared to those presented in Table 3. The coefficients on the level of innovation in the sending ZIP code are also significantly downward biased with respect to the results presented in Table 3.

Table 6: IV Poisson Models for Diffusion of Innovation

Distance & Periods	in levels	interacted	in levels	interacted
	(1)	(2)	(3)	(4)
First Diffusion (1/0)	2.9065* (1.3576)	3.5714*** (0.8869)	2.2281 (1.6865)	2.8985** (1.0242)
$\ln \text{Net Diffusion}_{r,t}$	2.2968* (0.9123)	2.5902*** (0.3974)	2.2130* (0.9844)	2.5803*** (0.4060)
$\ln \text{Distance}_{s,r,t}$	-0.4192*** (0.1108)	-0.2482* (0.1259)	-0.4137*** (0.1070)	-0.2369* (0.1207)
$\ln \text{Innovation}_{s,t}$	1.4029*** (0.3773)	0.8033* (0.3316)	1.4344*** (0.3833)	0.8093* (0.3290)
$\ln \text{Years linked}_{s,r,t}$			-1.0992* (0.4381)	-0.8859* (0.3636)
Constant	-73.3123 (43.1934)	-72.5021 (51.1024)	-1.1e + 02 (72.1243)	-97.6554 (58.9651)
Observations	268917	268917	268917	268917
Marginal Effects				
$\ln \text{Distance}_{\text{pre } 1996}$	-0.0167** (0.0061)	-0.0098 (0.0059)	-0.0165** (0.0059)	-0.0093 (0.0056)
$\ln \text{Distance}_{1996-2005}$	-0.0270*** (0.0071)	-0.0370** (0.0118)	-0.0260*** (0.0065)	-0.0364** (0.0119)
$\ln \text{Distance}_{\text{post } 2005}$	-0.0293** (0.0104)	-0.0159 (0.0098)	-0.0288** (0.0099)	-0.0161 (0.0100)

<sup>1</sup> Robust standard errors clustered at combination of sending and receiving census tracts: +  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

<sup>2</sup> All models include fixed effects for the sending census tract. Models also include dummies for same state and same city as well as a time trend variable.

<sup>3</sup> Instruments:  $\ln \text{Innovation}_{s,t-1}$ ,  $\ln \text{Innovation}_{r,t-1}$ ,  $\ln \text{Net Diffusion}_{r,t-1}$ ,  $\ln \text{Net Diffusion}_{s,t-1}$

Tables 5 and 6 are provided to address the question whether the results presented in Table 3 might be

affected by changes in ZIP codes during the period we study. We have no reason to expect that ZIP code changes will be endogenous to the diffusion of innovations in the United States, but it is conceivable that they might be.

Table 5 shows that dummy variables capturing cases in which the same location (longitude, latitude) is present under at least two different ZIP codes in our data are not significant or just significant at the 5% significance level in our data. Including these dummy variables also does not change the coefficients on the main variables of interest in a significant way. This indicates that ZIP code changes are not likely to affect our findings.

Table 6 is based on a different basis for local aggregation from the remaining results in the paper. Here we aggregate using US census tracts, which are made available through [Manson et al. \(2019\)](#). This results in a coarser set of locations and a smaller sample. The advantage of using this basis for regional aggregation is that census tracts are consistently defined for the entire sample period. The disadvantage is that we cannot use all of the same fixed effects controls as in our main models - the Poisson estimators do not converge when we do this.

Qualitatively the results we obtain from this aggregation are quite similar to our main findings: distance has a significant negative effect on diffusion and both absorptive capacity and innovation have positive effects on the level of diffusion. As we are unable to control for time fixed effects and Nice class fixed effects we do not place too much emphasis on these results.