The Impact of Broadband Internet on Invention: Evidence from the U.K.

By Sam Arts and Mattia Nardotto and Thomas Schaper and Jesse Wursten*

Draft: July 31, 2021

Prior knowledge is a crucial input for the creation of new technologies. In this paper, we investigate the impact of broadband internet diffusion on knowledge sourcing and patenting activity. Exploiting geographical variation in ADSL activation in the U.K. between 2000-2007, we find that broadband access increased the number of yearly filed patents by 10-55%. This increase was arguably driven by a better access to geographically distant knowledge, and particularly more recent foreign knowledge, as evidenced by an increase in the share of foreign patent citations, and a 2-3 year decline in the average age of these foreign cited patents.

JEL: O31, O32, O33, O52

Keywords: cumulative innovation, information technology, patents

In the endogenous growth theory literature, cumulative innovation constitutes a fundamental underpinning of economic progress (Romer, 1990; Aghion and Howitt, 1992; Jones, 1995; Weitzman, 1998). Prior knowledge and ideas are the main input in the production of new technological inventions. Edison's light bulb, Mullis' polymerase chain reaction to multiply DNA, or Google's PageRank are just a few famous examples. Larry Page, for instance, heavily relied on prior art ranking scientific publications in order to create Google's

* KU Leuven, Naamsestraat 69, 3000 Leuven, Belgium. Arts: sam.arts@kuleuven.be Nardotto: mattia.nardotto@kuleuven.be Schaper: thomas.schaper@kuleuven.be Wursten: jesse.wursten@kuleuven.be. Schaper and Wursten gratefully acknowledge financial support of the Research Foundation Flanders (FWO) under grant number 11B5918N and 1125818N. We thank Geert Dhaene, Reinhilde Veugelers, and seminar participants at KU Leuven for useful comments and suggestions.

search algorithm to rank web pages. In this regard, scholars have argued that technological progress and innovation heavily depend on the ability and cost to access relevant information (Jones, 2003; Mokyr, 2005).

Different large scale surveys across Europe and the U.S. suggest that technical knowledge from patent documents represents the most important source of codified information for inventors, next to scientific publications and tacit knowledge from interpersonal networks (Cohen, Nelson and Walsh, 2000; Jaffe, Trajtenberg and Fogarty, 2000; Giuri et al., 2007; Devlin, 2009; Sampat, 2010; Ouelette, 2012). While the patent system gives private incentives to incur the cost of innovation, the social value of granting patents is also contingent on the extent to which the publication of patent documents triggers knowledge spillovers and cumulative innovation (Machlup and Penrose, 1950; Scotchmer and Green, 1990; Scotchmer, 1991). A large body of prior work confirms that patent disclosure stimulates knowledge diffusion and technological progress (Graham and Hegde, 2015; Hegde and Luo, 2018; Furman, Nagler and Watzinger, 2018; Baruffaldi and Simeth, 2020; Hegde, Herkenhoff and Zhu, 2020; Lück et al., 2019; de Rassenfosse, Pellegrino and Raiteri, 2020). However, there is very little research and evidence on how inventors can access and search technical information in patent documents, and whether reducing the access cost triggers knowledge diffusion and innovation. One noticeable exception is Furman, Nagler and Watzinger (2018), who document a significant increase in patenting in the 48 regions who opened a new U.S. patent library, i.e. a central repository of physical patent documents, between 1975 and 1997.

In this paper, we contribute to this line of research by studying how the arrival of ADSL broadband internet to the United Kingdom in the early 2000s facilitated access to distant technical knowledge and led to a corresponding increase in innovation. Information technologies, like the printing press, the telephone, or computer networks, provide efficient means of storing and disseminating information (e.g. Dittmar, 2011). ADSL broadband, the first generation of high-speed and 'always-on' internet, entailed an exponential improvement in download and upload rates compared to basic cable internet. Broadband allowed inventors to efficiently and timely access codified technical information from the world wide web for the first time, in particular information stored in full-text online patent databases of both major patent offices (USPTO and EPO) and commercial providers. This made it much easier and less costly to access and search technical knowledge, particularly knowledge stemming from geographically distant regions and outside inventors' own social network and firm.

Linking proprietary, high-resolution data on the UK telecommunications market to public patent records, we evaluate how the switch from dial-up to broadband internet impacted knowledge sourcing and inventive activity in Great Britain. To do so, we exploit geographical discontinuities in activation status among 2,029 local exchange (LE) areas due to gradual delays in market entry of high-speed internet service providers (ISP) between 2000 and 2007. The roll-out of broadband happened through the upgrading of existing telephony nodes, called local exchanges, which had to be upgraded individually. Admittedly, the choice of market entry by ISPs in a LE is likely driven by firm establishments and local inventive workforce, even within narrowly defined areas. Therefore, we supplement our regional fixed effects estimations with an instrumental variables approach (IV) based on within region variation of weather conditions and areal demand mismatch between firm and local exchange positions that exogenously shift the likelihood of ISP entry, but are unrelated to firm/inventor establishment.

The launch of broadband in the UK provides a particularly well-suited research setting. This is because the roll-out of the infrastructure was not instantaneous, due to demand-side and supply-side factors, thus generating variation in internet availability between areas, even within small and homogeneous regions. The main causes of these initial delays in broadband expansion were due on the one hand to the uncertainty faced by the ISPs about the demand of internet services from private households, and on the other hand to the technical difficulties that the nation-wide roll-out of a new series of technologies naturally contemplated. In our empirical analysis we exploit these factors to develop a set of exogenous shifters that we use in an instrumental variable setting, to study the effect of internet access on innovation. The first instrument hinges on the negative impact of adverse weather conditions on broadband take up and on service quality (Gavazza, Nardotto and Valletti, 2019). Weather conditions have been shown to have an important weight in ISPs entry decision, as broadband equipment were (and still are) particularly sensitive to adverse weather, which strongly impacted maintenance costs and service quality. Thus, this class of instruments uses variation in weather conditions, namely rainfall, between local exchange areas to predict broadband activation in the following year. The identifying assumption behind our approach is that differences in weather conditions between LE areas *within NUTS3 regions* are orthogonal to the inventive firms' and inventors' characteristics established in these areas.[1] The second class of instruments exploits potential mismatches between demographic - hence demand relevant - characteristics of local exchange catchment areas and (partially) overlapping Census 2001 Middle Layer Super Output Areas (MSOA).[2] The intuition behind this approach is that firms and inventors base their establishment decisions *within NUTS3 regions* on current population characteristics of MSOAs, while they are unaware of the extension of the (historical) catchment areas of the nodes of the telecommunications network, which are relevant to the internet service providers.

As main result, we find that the activation of broadband increased yearly patenting in a LE area between 10-55%, and forward citation-weighted patenting by %, depending on the specification employed. The latter finding suggests broadband particularly resulted in higher quality and more impactful inventions. We relate these extensive margin effects to the reduced costs of accessing and searching technical knowledge from abroad. In particular, we show that broadband access increased the likelihood to cite foreign technical prior art by 4-14%. Moreover, broadband enabled inventors to search and build on more recent foreign technical prior art, suggesting it allows inventors to keep up to date on the knowledge frontier and state of the art in given field, particularly if new developments originate in remote areas. Apart from sourcing (recent) foreign knowledge, we find no significant

[1] NUTS3 regions correspond to territorial subdivisions of between 150,000 and 800,000 inhabitants. In total there are 174 NUTS-3 regions in the U.K, 167 of which generate patents in our data.

[2] An MSOA is a statistical area covering a population of between 5,000 7,200 inhabitants. There are 2,029 LEs and 3,374 patent generating MSOAs in our data.

effect of broadband on sourcing distant knowledge from outside inventors' own social network or firm, or from different technology fields. Furthermore, besides affecting knowledge sourcing, high-seed internet might also influence technological progress and innovation by changing the organization of R&D departments and teams, e.g. through a reduction of communication costs and resource-sharing across multiple branches. However, we find no significant increase in team size, geographically dispersed teams, cross-disciplinary teams, or new collaborations among inventors. These findings strengthen our belief that a better and timely access to foreign technical knowledge triggered the increase in innovation. This finding is closely in line with (Iaria, Schwarz and Waldinger, 2018), who show that World War I caused a significant decline in international scientific references in papers, and consequently reduced scientific progress and output.

To further strengthen the causal claim that early broadband increased patenting through a reduction of access costs to (recent) foreign knowledge, we conduct a number of additional tests. First, broadband could also affect innovation in LE areas through the opportunity it creates, for instance by attracting new companies and inventors.[3] Reassuringly, we find that our results hold when we restrict our sample to firms established in a region before the arrival of broadband, ruling out that our effects are driven by the location choices of firms. Moreover, the introduction of broadband did not attract a significant number of new inventors to a LE area. Second, we show that broadband access in LEs did not significantly affect public sector patenting (i.e. by universities, government research labs or hospitals). Indeed, prior literature highlights that these institutions were pre-embedded into different types of information networks (e.g. BitNet, ARPAnet, physical libraries), making them arguably exposed to a much lesser shock when broadband arrived.[4] Third, we find no effect of broadband on the characteristics of examiner-given citations. Indeed USPTO patent examiners work from the patent office in Virginia in the US and should not be affected by the arrival of broadband in the UK. Fourth, we observe no change

[3]see, e.g. Cheyre, Klepper and Veloso (2015); Akcigit, Baslandze and Stantcheva (2016) regarding determinants of inventor mobility.

[4]for a discussion on this, see e.g. MacKie-Mason and Varian (1996); Spear (1996)

in the overall number of prior art references per patent, allowing us to exclude that the effect of broadband was merely one of patent attorneys or inventors becoming more aware of which patents they would need to cite for legal reasons, without the actual manifestation of knowledge spillovers. Finally, we do not see an increase in references to scientific publication in patents. While patent documents became publicly available online in 1999, accessing scientific publications still required expensive subscriptions to publishers or library access (e.g. through academic affiliations).

However, a series of historical placebo tests reveal significant pre-trends and higher historical patenting rates of LE areas within our chosen exclusion restrictions. This warrants caution for the causal interpretability of results from our (current) IV-estimation. We discuss motives and consequences of this limitation and discuss a series of potential remedies and alternative approaches devised in the attempt to re-establish causal identification.

Our paper has antecedents in the empirical literature regarding the importance of access to existing knowledge for subsequent scientific and technological progress (Moser and Voena, 2012; Williams, 2013; Iaria, Schwarz and Waldinger, 2018). In particular, we contribute to the debate about how access costs to prior information influence research productivity and shape trajectories of scientific and technological inquiry (Furman and Stern (2011); Bryan and Ozcan (2016); Biasi and Moser (2018)). Our findings suggest that the elasticity of the cost of access on technological progress and innovation is significant and large, even in the 21st century, and that it has real economic effects. Moreover, we provide evidence in support of the argument that access costs constrain the diffusion of technological knowledge embodied in patent documents. In this regard, our results also closely relate to recent empirical work showing how the disclosure of patent documents fosters knowledge spillovers and cumulative innovation (Graham and Hegde, 2015; Hegde and Luo, 2018; Furman, Nagler and Watzinger, 2018; Baruffaldi and Simeth, 2020; Hegde, Herkenhoff and Zhu, 2020; Lück et al., 2019; Zheng and Wang, 2020; de Rassenfosse, Pellegrino and Raiteri, 2020). While our data does not allow us to make inferences about whether inventors acquired technical information from patent documents through online patent databases or other channels, we causally show that broadband access enabled inventors to become aware of and built on technical prior art they would have otherwise not.

Our findings further contribute to the literature on the effects of ICT adoption and internet on firm performance (Forman and van Zeebroeck, 2012; Bertschek, Cerquera and Klein, 2013; Dhyne et al., 2018; Forman and van Zeebroeck, 2019). These studies show internet adoption impacts intra-firm knowledge diffusion, R&D collaboration and productivity. Other related work has documented the positive effects of BitNet, a predecessor of cable internet, on academic research collaborations and productivity (Agrawal and Goldfarb, 2008; Ding et al., 2010). Finally, given the comprehensive, country-wide evidence of our study, our results also speak to the wider literature on the impacts of information technologies on aggregate growth, employment and productivity (e.g., Dewan and Kraemer, 2000; Black and Lynch, 2001; Stiroh, 2002; Dittmar, 2011), and of broadband infrastructure in particular (e.g., Czernich et al., 2011; Hjort and Poulsen, 2019).

I. The UK broadband roll-out and the search for technical prior art

In this section, we outline the institutional context of our investigation by discussing the evolution of the UK broadband market, followed by a depiction of how inventors can search technical information from patent documents, before and after the arrival of broadband internet.

A. The UK broadband market

At the beginning of 2020s, thirty years after the advent of the internet, the market for internet services relies on a mix of technologies and networks – xDSL, Cable, Fiber to the Home, or Mobile internet – and it hosts a good number of Telcom companies competing for internet users. At its infancy in the 1990s however, internet services were provided only through one technology and one network: the copper lines of the telephone network. In the UK, this network was owned and managed by British Telcom (BT), the formerly state-owned telecommunication company which in the 1930s set up the network of nodes (called Local Exchanges, LEs) and links that connect each household to the rest of the country and, through undersea connections, to the rest of the world.

The telephone line proved to be a viable option for the transmission of a digital signal, thanks to an upgrade to the old-fashioned telephone connection. This network, although it was not meant for that purpose, had the advantage of being already deployed and covering virtually the whole UK (due to the universal coverage obligation). Until the end of the 1990s tough, i.e. until the moment when the first family of Digital Subscriber Line (DSL) technologies were introduced, internet users could only connect to the internet trough dial-up (modem) connections, which at most could reach a (download) speed of 64 kbit/s. The introduction of DSL marked the start of a new era due to its large improvements in data transmission speed. With the first generation of DSL technologies introduced in the early 2000s in the UK, the speed of internet connection increased up to the (theoretical) speed of 8 Mbit/s, i.e. by a factor of 125.[5]

This fundamental technological step happened right after the release of online patent databases by the major patent offices and commercial providers, which enabled inventors to efficiently and timely search and download the full text of virtually the entire patent record of the major patent offices (see further, section I.B). However, internet access at sufficient speed however strongly depended on two conditions that had to be met at the level of the LE to which the inventor was connected: i) The LE being upgraded by BT to host ADSL connections; ii) The LE receiving the investments of Internet Services Providers (ISPs) to offer such high-speed ADSL connections. In order to understand the relevance of these two conditions, it is worth reconstructing how the broadband market evolved in the UK during the 1990s and the first decade of the 2000s. When the internet era started, the privatization process of BT was almost completed.[6] Before its privatization, BT was vertically integrated, meaning that it was both managing the telephone network, and selling

[5]As documented in Nardotto, Valletti and Verboven (2015), and Ahlfeldt, Koutroumpis and Valletti (2017), and several other studies, actual speed was much lower than the theoretical speed that could be reached by the technology due to factors, such as the physical distance to the node, the weather, and other conditions, which determine a deterioration of the quality of the signal and thus a drastic decrease in the actual speed of the connection.

[6]The company started its process of privatization in the mid-1980s, and concluded it in the mid-1990s.

telephone services to the final users. This was a source of concern for policy makers once the social and economic potential of the internet became clear. The reason was simply that BT's market power in the telephone services market could be transferred to the new market of internet services. This led to various actions to vertically separate the two levels of the market and to facilitate entry of ISPs in the downstream market of internet services.[7] BT was separated into two companies, Openreach (in charge of managing the network) and BT Retail (investing in the LEs and selling internet access to final users), and a set of *open access* policies were introduced to boost entry of other ISPs in the downstream market. The market for broadband services could then profit from an investment-friendly regulatory framework where BT, and subsequently Openreach, was mandated to quickly enable the LEs to host ADSL investments. BT charged fees in exchange for access to the network and ISPs could freely enter the downstream market by investing in the necessary technology at the level of a LE in order to serve internet users, both private and corporate.

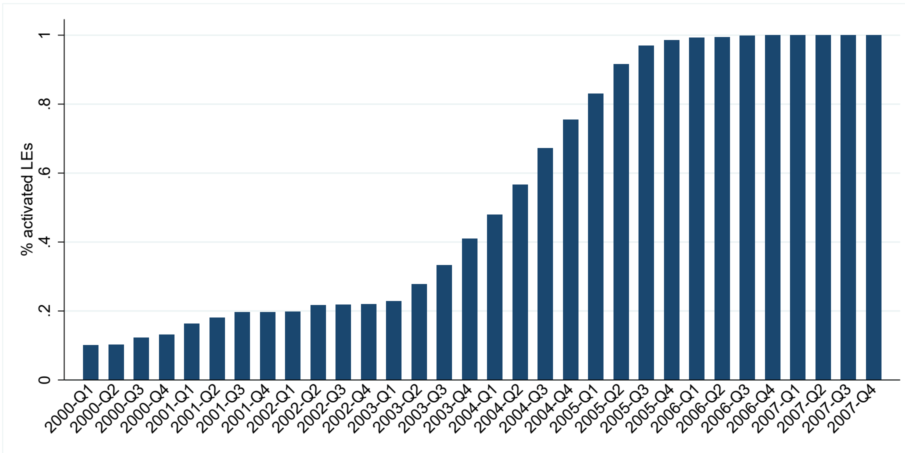
The combination of investment-friendly regulation and growing demand for internet access determined the rapid diffusion of the internet. Broadband penetration, measured as the share of households with a broadband connection, rose from less than 10% in 2000, to more than 80% in 2010. However, the market did not grow at a constant rate, with modest growth rates until 2005, to effectively take-off in the second part of the 2000s when private internet users massively wanted to be online.[8] In the end of the 1990s and the first part of the 2000s the process of upstream LE activation by BT was much ahead of the actual downstream investment rate of ISPs because of the high installation costs[9] and a relatively weak demand for internet services. This slow take up of investments by ISPs is reflected in Figure 1 which reports the share of LEs which received ADSL investments of at least

[7]The actions taken in the UK were in line with those taken in the rest of EU countries, which had to confront a very similar market structure characterized by the presence of old incumbents which were formerly state-owned companies. A relevant role in shaping the institutional framework was played by the European Commission through its EC Directives.

[8]Private users in the second part of the 2000s could finally enjoy a combination of newer/better ADSL technologies, and the arrival of new content such as Youtube (founded in 2005), Facebook (introduced in UK in 2007) and other successful platforms.

[9]According to Nardotto, Valletti and Verboven (2015), who examine the period between 2005 and 2009 the cost to enable a LEs halved, thanks to technological improvements and learning. This figure is probably much larger if one considers the costs in the early 2000s.

Figure 1. Share of LEs with broadband access by quarter



one ISP between 2000 and 2007. The roll-out of the technology started at a slow pace and between the end of the 1990s and mid 2003 only 20% of LEs provided ADSL broadband connections to private users and firms. In 2004 the diffusion finally took-off, and in the following 3 years the market was fully covered. This slow diffusion of internet access due to ISPs' investment decisions is at the core of our identification strategy, which exploits the timing of the introduction of broadband services between different LE areas (see section III.A).

B. The search for technical prior art online

Existing knowledge and prior art are a fundamental input for the creation of new technologies. Industrial R&D in particular relies on external knowledge spillovers, i.e. knowledge stemming from outside the firm and team of inventors (e.g., Jaffe, 1986; Bloom, Schankerman and Van Reenen, 2013). During the inventive search process, inventors can rely on codified knowledge from predominantly scientific publications and patent documents, or tacit knowledge stemming from personal interactions and social networks. Prior empirical work shows that interpersonal networks of inventors indeed foster the diffusion of knowledge, and that these networks explain the frequently observed localization of knowledge flows within firms and geographic regions Jaffe, Trajtenberg and Henderson (1993); Verspagen and Schoenmakers (2004); Singh (2005); Singh and Marx (2013). Despite the importance of external knowledge for industrial R&D, there is less evidence on the channels through which inventors access and search codified knowledge outside their own social network, firm, and geographic region. Large scale inventor surveys in Europe and the U.S. point out that patent documents are the most important source of codified knowledge next to scientific publications and conference proceedings (Giuri et al., 2007; Cohen, Nelson and Walsh, 2000; Jaffe, Trajtenberg and Fogarty, 2000).[10] More recent studies confirm that the public disclosure of patent documents triggers external knowledge spillovers, technological progress, and cumulative innovation (Graham and Hegde, 2015; Hegde and Luo, 2018; Furman, Nagler and Watzinger, 2018; Baruffaldi and Simeth, 2020; Hegde, Herkenhoff and Zhu, 2020; Lück et al., 2019; de Rassenfosse, Pellegrino and Raiteri, 2020)

Despite the importance of open and early access to patent documents, corporate inventors initially faced considerable hurdles and costs to access and search technical information from patents. Before the internet, the default was to keep track of the state of the art and new developments in a field by screening periodical bulletins containing titles and abstracts of newly granted patents, and subsequently ordering individual copies of selected patent documents from the central patent office via mail or fax. This search process was inefficient, slow, and costly. For instance, for UK-based inventors, the fee for regular mail delivery for a single copy of one patent document from the USPTO ranged between $5-19 current USD in 1997, with a delivery time of 4-6 weeks.[11] Fax delivery was much faster but also more expensive with fees ranging between $41-60 current USD for a single patent copy international delivery. [12] Besides ordering individual copies by mail and fax, people could also consult physical patent document repositories in patent libraries or information centres

[10] In particular legal scholars have often questioned the usefulness of patent documents as sources for information for inventors (e.g. Devlin (2009); Ouelette (2012); Lemley (2012). Main concerns raised in this regard relate to the little incentives inventors have to disclosure actual valuable information in their own patents and to consult other patents, due to the risk of committing "willful infringement".

[11] We determine inflation adjusted prices in current $USD based on Consumer Price Index (CPI) data from the U.S. Labor Department's Bureau of Labor Statistics (cumulative inflation rate through February 2020).

[12] Source: Historical website of the USPTO, accessed Feb 12 1997.

of the UK patent office located in large metropolitan areas across the country.[13] Notice that there were fewer than 10 of such patent libraries across the UK in 1995.[14] Finally, a number of university libraries in the UK were also connected to the CASSIS system through the ARPAnet, enabling access to the electronic repository of patent documents from the USPTO MacKie-Mason and Varian (1996); Furman, Nagler and Watzinger (2018).

Dial-up internet made many things easier for inventors in the mid-1990s.. Commercial service providers, such as Questel-Orbit and Derwent, were the first to realize the potential of the world wide web to offer online patent information and sell electronic full text copies of patents at comparatively lower prices.[15] The major patent offices started their free online databases with some delay. In 1996, the USPTO was the first patent office to offer a comprehensive online catalogue with bibliographic front cover information and abstracts for over 2 million patents.[16] This was followed by an update in late 1998, making the full text and images of all US patents available online. In 1999, the EPO launched its online platform *Espacenet*, which included full text information for all patents worldwide.[17]

Despite the online availability of patent information from commercial providers and patent offices, the critical problem with searching and retrieving patent documents was the limited bandwidth and slow download rates of 56k-modems (see section I.A). Searching for, selecting, and opening a full text patent document from a dial-up connection meant, in the best case, several minutes of waiting time until all pages and drawings were downloaded. In many instances, though, web pages were simply not reachable, as the early infrastructure of the net was not capable of handling the torrent of user requests during peak times (Spear, 1996). Taken together, accessing and searching patent information online remained a tedious, costly, and at best complementary tool in the late-1990s.

Broadband internet dramatically decreased the costs of accessing information online

[13] The PatLib program was an initiative by the EPO and national patent offices targeted at providing access to patents and trademarks to individual inventors and businesses.

[14] Source: WIPO 2015.

[15] For an extensive discussion of patent information available online to UK inventors in 1996, see Spear (1996).

[16] This was preceded by a pilot project offering full text and image access to a limited number of HIV/AIDS related patents in 1994 on the USPTO website (Source: USPTO).

[17] Sources: Press releases/ historical archives of the USPTO and EPO websites.

and improved the speed and easy of use. Patent documents, technical reports, and scientific publications could now be browsed and downloaded within seconds. Moreover, DSL-technologies enabled parallel use of telephone, fax and internet for the first time. This 'always-on' feature of broadband enabled internet search to be fully integrated as a simultaneous functionality into many research and development activities.

II. The Data

In this section we present the data sources that we employ in the empirical analyses: i) data on the broadband market; ii) patent data; and iii) socio-demographic data.

A. Broadband market data.

Data on the evolution of the broadband market, which we discussed in Section I.A, have been provided by Ofcom, the UK regulation and competition authority for the broadcasting, telecommunications and postal industries. The data includes the year and month of arrival of ADSL for each of the approximately 5,500 LEs in the UK, covering the countries of England, Wales and Scotland. The data also contains the exact geographical catchment area covered by each LE. We use this information to determine for each inventor the LE she/he is connected to and the start date at which she/he could access the internet at broadband speed. Notice that despite the high granularity of our data on micro-market entry of broadband service providers in a LE, we do not observe actual broadband adoption at the inventor level.

B. Data on patenting activity

We rely on public patent records to quantify the yearly inventive output in a LE and to characterize the knowledge and prior art on which the inventors relied during the inventive search process.

We start by collecting information on all patents granted by the USPTO since 1976 from *PATSTAT 2018* (spring edition). We rely on data from the USPTO because US patents provide the most comprehensive coverage of all inventive activity worldwide. The majority of all patented inventions (approximately 60%) has a USPTO patent family member. Moreover, UK firms predominantly file patents at the USPTO rather than at the UK or European Patent Office Griffith, Harrison and Van Reenen (2006). Therefore, US patents provide a better measure of the inventive output of UK based firms. Finally, the USPTO provides the best and most reliable coverage of information on citing and cited patents' locations, that is indispensable for our analyses.

First, we collect information on the geographic location of assignees and inventors for all patents granted by the USPTO.[18] From the bulk of US patents, we restrict the sample to corporate patents having the assignee and at least one inventor located in the United Kingdom. [19] [20] Because the USPTO does not provide fine-grained address information for all inventors and assignees, we replenish coarse geo-locations with high-resolution address strings sourced from international patent family members, i.e. patents issued by other national and regional patent offices covering the same invention as the US patent Morrison, Riccaboni and Pammolli (2017); de Rassenfosse, Kozak and Seliger (2019). First, we rely on European Patent Office (EPO) family members, which have the highest-precision and most comprehensive address string coverage. % of all US patents in our sample have an EP family member. Second, for the remaining US patents without an EP family member (%), we obtained additional information on applicant and inventor addresses for UK patent family members directly from the United Kingdom Intellectual Property Office (IPO).[21] Next, for the few patents in our sample without an EP or IPO family member (% of our sample), we rely on address information from patent family members at different patent offices, obtained from de Rassenfosse, Kozak and Seliger (2019)[22] We drop % of the patents in our sample for which we have no fine-grained location information. Finally, we geo-disambiguate all

[18]Note that, while a large body of work exists in the literature using geo-locations of US inventors, only less than 5% of all USPTO patents actually contain fine-grained address information of inventors (and applicants).

[19]We exclude public sector patents, i.e. patents filed by universities, public hospitals, government or military institutions, as these are typically embedded in different types of knowledge networks (e.g.BitNet) and inventive search is conducted under different boundary conditions (public funding, non-profit orientation,...)

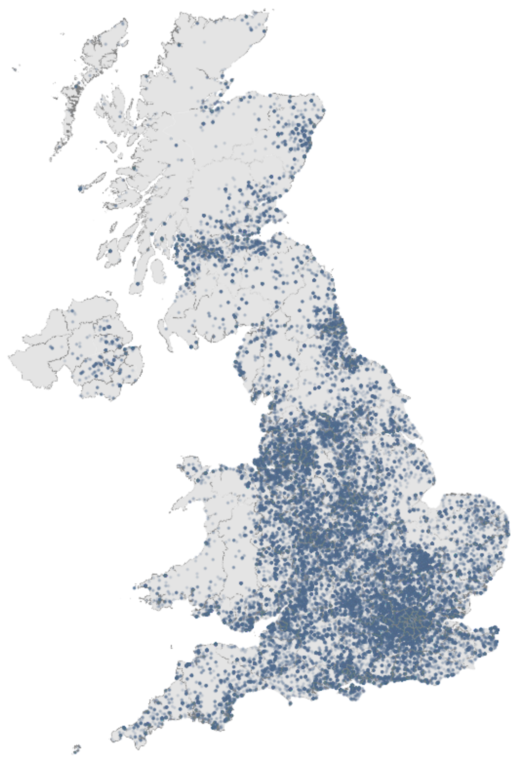
[20]We determine assignee types based on the PSN sector allocation in PATSTAT.

[21]We thank the IPO for providing us timely with the requested data sets and very useful further explanations.

[22]We thank Florian Seliger for providing us with the input data.

raw address strings in our data using the *HERE Developer* API, a commercial mapping and location data service.[23] Figure 2 provides an overview of the dispersion of inventive activity across the UK between 1975 and 2010. Historically, patenting appears to be clustered geographically, and concentrated in the south and south-east of Great Britain around the metropolitan and industrial centers of London, Liverpool, Manchester and Birmingham. Clusters of inventive activity are also visible in the north of England, between Newcastle and Middlesborough, as well as around Glasgow and Edinburgh in Scotland.

Figure 2. Geolocation of inventors across Great Britain 1975-2010



In a next step, we are interested in determining whether or not the inventors of a patent had access to ADSL broadband internet during the inventive search process. To determine broadband activation status, we use assignee address information because 1) inventor ad-

[23] API interface available at: https://developer.here.com/

dresses usually correspond to the private home rather than the workplace location (Pavitt and Patel, 1998; Griliches, 1990; Verspagen and Schoenmakers, 2004), because the workplace address most likely reflects the location of inventive activity, and because companies were typically faster to adopt broadband compared to private households. However, prior literature has warranted caution in using addresses of firms listed as assignee on a patent, as these might reflect the address of firm headquarters or the address of the legal entity that owns a specific patent, e.g. the parent firm holding (Jaffe, Trajtenberg and Henderson, 1993; Singh, 2008; Singh and Marx, 2013). To increase the likelihood that the assignee address corresponds to the actual workplace of the inventors, we restrict our sample to patents for which at least 75% of the inventor addresses are within a 75km radius around the assignee address, which we consider as an upper bound for plausible commuting distances. More than 90% of all patents in our sample fall within this definition, suggesting that the address of UK firms listed on US patents are likely to correspond to the workplace of the inventors. Since we only dispose of broadband activation information for UK addresses, to further avoid measurement error, we restrict our sample (in our main specification) to only those patents having all inventors located in the UK, i.e. excluding international collaborations. Notice that our results are robust for the subset of patents with 100% of inventors within the same 75km radius around the assignee address, as well as for 50km and 25km radii.[24] In case a patent is assigned to two assignee firms (less than 10% of patents), we consider only those assignee addresses for which the 75/75 definition is valid.

C. Census data

We collect additional information on the socio-demographic characteristics of different subsets of the population from the UK Census. The UK national census takes place during the first year of each decade. Given that our analysis spans the period between 2000 and 2007, we consider the Census of 2001. We collect both standard socio-demographic information – such as age structure, education, and unemployment rate – which will be used

[24]From our sample of patents for which 75% of the inventor addresses are located within 75km from the assignee address, 95% also fulfill the stricter 50km, and 83% the 25km criterion.

as controls, and variables – such as population density – which will be used to construct instrumental variables, as we discuss in Section III.A. To reconstruct demographic characteristics at the level of the LE, we downloaded information at the Output Area (OA) level, which is the smallest geographical unit at which the UK Census is available.[25]

D. Descriptive statistics

Table 1 provides descriptive statics for our final sample of 6,961 patents, which represents about half of all corporate USPTO patents with all assignee and inventors located in the UK during the same time window. To evaluate how biting the constraint of having high-resolution address information is for how generalizable our results are for the underlying population, we compare characteristics of in-sample to not in-sample patents in Table 1. The distributions across most patenting indicators are heavily skewed. The average patent in our sample lists 2 inventors, includes 22.01 prior art references to patents and 3.12 references to scientific publications. Over a period of 10 years, patents receive on average 18.24 citations by follow-up inventions (patent families). About 2% of patents are assigned to multiple corporate or institutional assignees.

The comparison with patents not in our sample reveals significant differences with regards to the distribution of technological impact (10-year forward citations), the number of patent references and, in particular, the number of 5-year patenting activity by the same corporate assignees. The median firm in our sample files 5 patents in 5 years, compared to 8 patents filed by not in-sample firms, for mean values this differences is about twice as large. We interpret this as a sign that large multinational firms are possibly underrepresented in our data, as they are more likely to not file patents directly through the national UK IPO, and less likely to assign patents to inventor workplace addresses. The remaining indicators show roughly comparable scores.

Table 2 provides yearly summary statistics of inventive activity at the LE area level.

[25]There are approximately 216,000 OAs in England, Wales and Scotland in the 2001 Census. The minimum population size for an OA was 40 resident households, but the vast majority, accounting for 79.6% of the total, lies between 110 and 139 households. LEs instead are 5,500 thus a LE contains on average almost 40 OAs. As the boarders of the two geographies do not perfectly match, we input the demographics of the areas that overlap with a boarder pro-quota based on the proportion of the overlapping area.

Table 1. Patent level summary statistics

	in-sample				not in-sample			
	p50	Mean	SD	Max	p50	Mean	SD	Max
Citations [10y]	11	18.24	24.03	281	12	19.86	30.53	502
Patent References	15	22.01	21.43	330	13	18.69	22.62	589
Scientific References	0	3.12	13.58	363	0	2.68	12.19	363
Inventors	2	2.06	1.43	16	2	2.14	1.52	17
Co-Patent [01]	0	0.02	0.15	1	0	0.04	0.20	1
Assignee 5y Patents	5	76.46	362.74	5.1k	8	141.31	453.83	5.6k
Assignee 5y Citations	36	721.68	2.9k	65k	80	1.6k	4.7k	102k
N of Patents	6,961				6,703			

Notes: In-sample patents are patents with high-resolution address information allowing to map them accurately to Local Exchange catchment areas. Patent data are sourced from PATSTAT, USPTO, and BvD Orbis (assignees). Additional applicants and inventors address information was obtained from the U.K. IPO and de Rassenfosse, Kozak and Seliger (2019).

We register patents originating from 2,029 LE areas during our sample window. As seen in Figure 2, patenting is not evenly distributed across nodes. While the most active LE generates 114 patents in one year, the average number of filings per LE is merely 0.55. Around 62% of patents are assigned to companies established in the region before the arrival of broadband. The amount of citation weighted patents per LE for pre- versus newly established applicants is proportional to this division. About half of the LE have no patents in the 5 years before the arrival of ADSL.

III. Empirical analysis and results

In this section we present our empirical analysis. We start by the identification strategy, and we discuss the challenges posed by the potential sources of endogeneity that are related to the roll-out of the ADSL infrastructure. We then show our findings and we present evidence that earlier internet access led to more patents, which also receive more citations. The main mechanisms appear to be more international and faster knowledge sourcing.

Table 2. Node (LE) level yearly summary statistics

	Mean	SD	Max
Patents	0.55	2.27	114
Citation weighted Patents [10y]	9.86	68.44	4.9k
Patents (pre-BB Firms)	0.34	1.91	114
Citation weighted Patents (pre-BB Firms)	5.99	47.55	2.5k
Public Patentee [01]	0.10	0.30	1
Pre-BB Patenting Node [01]	0.49	0.50	1
N of Nodes	2,029		

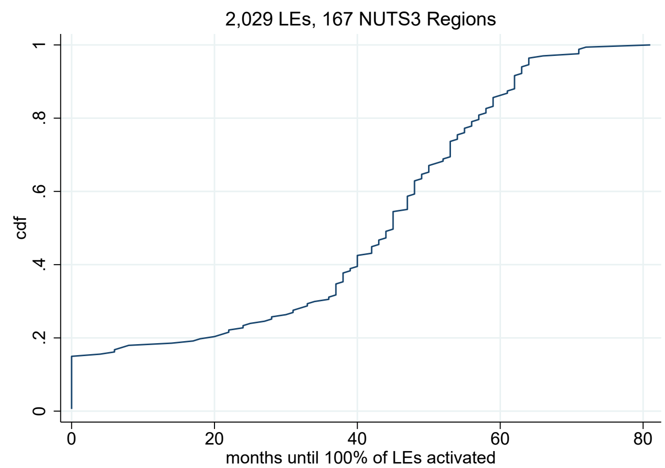
Notes: Nodes (LEs) are Local Exchange Points geographical catchment areas (N = 5,500) across the U.K.. Data on LEs and ADSL activation are sourced from Ofcom.

A. *Identification strategy*

The goal of this paper is to document the effect of the internet on patenting. To do so, our identification strategy exploits the variation in timing of introduction of broadband internet in local areas. As discussed in Section I.A, the roll-out of ADSL nationwide was not instantaneous, and both demand and cost factors affected the decisions of ISPs to invest in the LEs. Importantly, this process of progressive roll-out was also happening at the local level. Figure 3 reports the average time (in months) needed to reach different levels of ADSL coverage within the NUTS 3 regions, i.e. a similar exercise to the one in Figure 1 in Section I.A, but performed at the local level. As Figure 3 shows, the roll-out was not only a slow process from a nationwide perspective, but also within regions, thus generating the variation in timing of introduction that we exploit to estimate the effect of internet on patenting activity.

However, we cannot consider this roll-out process fully exogenous. In particular, the core endogeneity concern is that certain areas are more attractive to both ISPs and innovative firms. For example, we would expect both more patenting activity and earlier internet activity in the urbanized web around London than in the scenic vistas of Cornwall. We address this concern with two sets of instruments, which are based on two different mechanisms driving ISPs' investments decision, in a way that they are not correlated with firms'

Figure 3. Cumulative of *within NUTS3* ADSL Activation Lag across Great Britain



R&D or location decisions conditional on our controls. Before discussing the instruments, it is important to notice that the decision of an ISP to invest in a specific LE was not driven by the demand coming from the firms we are considering in this study. Instead, the main source of revenues, and thus the key determinant of an ISP's decision are the expected private internet connections, as they outnumbered by far the rest of the market. Thus, factors that affected the economic trade-off of serving the local internet markets can be exploited once we control for the characteristics that are attractive for firms and inventors in general.

The first set of instruments follows the identification strategy in Gavazza, Nardotto and Valletti (2019). Thus, we exploit weather conditions as a source of variation for the timing of investments in the LEs. In particular, we use the amount of rainfall in a LE, which affects both the cost of upgrading a LE for the ISPs and local demand for internet services, as more rain reduces service quality, as argued in great detail by Gavazza, Nardotto and Valletti (2019). They note that there is an 0.8 correlation between rainfall and network fault report volumes as "these conditions led to more underground structures flooding and more faults due to water ingress into failed joints and cables" (Ofcom, 2014). Moreover, adverse weather conditions lead to reduced reliability and speed of the internet service, in turn reducing local demand (Bouckaert, van Dijk and Verboven, 2010; Nardotto, Valletti and Verboven, 2015; Gavazza, Nardotto and Valletti, 2019).

We include NUTS3 fixed effects, therefore the exclusion restriction holds unless innovative firms are sensitive to rain conditions within NUTS3 regions. That would require two strong assumptions. First, that firms are aware of differences in rainfall within relative small regions and second that this is a meaningful determinant of their location choices. The first is implausible, as rainfall data have only recently become publicly available and are not trivial in their use. Moreover, the difference in rainfall within a NUTS3 region is sufficiently large to matter for the highly sensitive electrical equipment of ISPs, but not so large that you would notice it through anecdotal experience. In Figure **??** we show variation in rainfall for urban Birmingham and rural Sefton. The difference between the driest and wettest point is respectively 20% and 27%, large enough to affect the vulnerability of the ISP equipment, but not enough to be noticeable without access to detailed rain data.

Second, firms being aware of differences in rainfall would only affect our estimates if they base their location decisions on this information. Yet, there do not appear to be any reasons to assume mild differences in rainfall would matter to businesses. Instead, a survey by Keeble and Nachum (2001) suggests that firms mainly care about proximity to clients and talent and does not even mention weather conditions. As a placebo test, we examine whether rainfall predicts public sector patenting. Public institutions largely have their own internet networks and are thus not affected by ISP decisions, but if firms care about local rain conditions, it stands to reason that public institutions do too. However, as we will show in Table 6 we find a precisely estimated null effect, providing suggestive evidence that variations in rainfall within NUTS3 regions truly are exogenous.

The second set of instruments exploits the as-good-as-randomly defined catchment areas of the local exchanges which are determined by the British Telecom's telephone network as it was constructed in 1930 and are not used for any other purpose, and have not been changed since (Gavazza, Nardotto and Valletti, 2019). Identification is then driven exclusively by the mismatch between the area of relevance for ISPs and for firms. We use Figure 4 to illustrate this identification approach. In the figure, blue dots indicate the

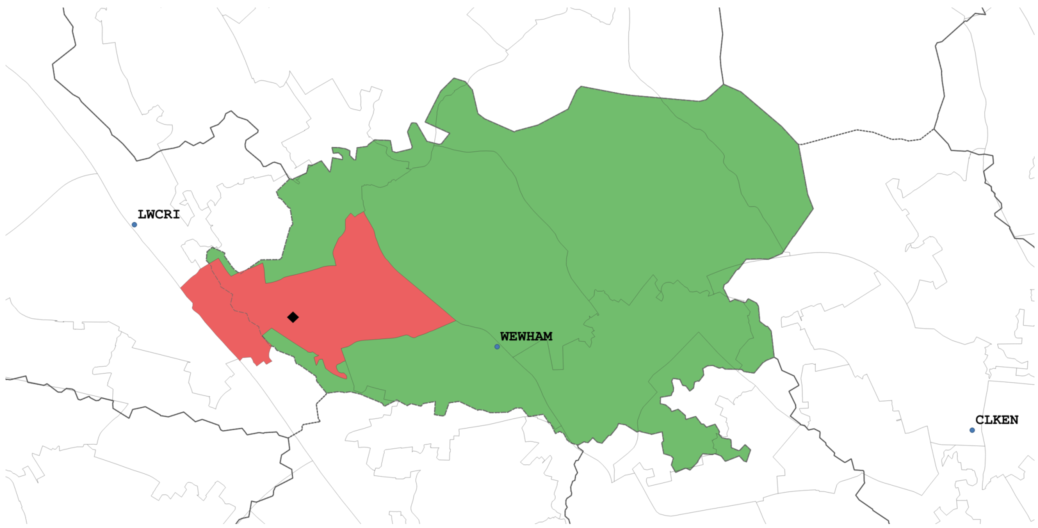


Figure 4. Example of LE catchment area and UK's MSOA. The dark solid lines are the LEs' boarders while light solid lines are the boarders of the MSOAs. LEs' locations are identified by the blue dots (in the figure we have the LEs named WEWHAM, LWCRI, and CLKEN). In this example, a firm located in the black diamond, is part of the red MSOA and it is served by the LE WEWHAM, whose catchment area is the one in green.

position LEs, while their catchment area is delimited by the dashed thick line. The solid gray lines are the boarders of the MSOAs. An MSOA comprises on average 6,500 families, which means, on average 15,000 inhabitants. MSOAs are census areas and are wide enough to capture well the characteristics of the section of the city or of the town where the firm and the inventor is operating. In the figure we have a fictitious firm located at the diamond. In our data we have exact firm's location so, for each firm, we know exactly the relevant MSOA, in red in our example, and the relevant LE, which we depict in green in our example. The core of the identification strategy relies on the mismatches between the characteristics of the MSOA, relevant for the firm, and those of the LE's catchment area, which are relevant for the ISPs. So, while controlling for both the characteristics of the MSOA and of the wider NUTS3 region, we use the demographics of the remaining part of the LE as instrument for the arrival of broadband investments. This instrument resembles the one proposed by Hausman (1996) and later applied by Gentzkow and Shapiro (2010), or Fan (2013) which hinge on the fact that the demographic characteristics of the neighboring areas of a given area i provide incentives for the supply-side to take actions (pricing in their case, investing in ours) which can affect also the area i. Clearly, this supply-side connection is present in our case, as the investment decisions taken by ISPs are based on the characteristics of the (entire) catchment area of the LE, but then they affect every building which is located on top of it, so all MSOAs that are covered by the LE.

One limitation of this method is that while we know exactly which area matters for the ISP, we do not have that information for firms. This is less of an issue than one might expect as the NUTS3 fixed effects control for characteristics of the broader region. Therefore, any endogeneity would have to arise from innovative firms basing their location decisions on some other subset of the NUTS3 region that overlaps more strongly with the (random) local exchange catchment area than with the carefully constructed MSOAs. Moreover, there would need to be a link between the number of innovative firms and the chosen demographics which is not a trivial assumption.

The instruments we use are population density, a set of age share variables and the total number of people, each measured within the catchment area.[26] We include the same variables as controls, but this time measured at the MSOA level. We also include NUTS3 fixed effects (MSOAs are smaller than NUTS3 regions).

B. *Inventive Activity by LEs*

Our first goal is to determine whether broadband access lead to increased patenting. The timing of this process is non-trivial. We know the filing date of each patent[27], but not necessarily when the main inventive search period took place. Moreover, we know when the first ISP entered the LE, but not when the firm actually buys broadband access, nor how

[26] Ages are split 0-14-19-29-44-59, with 60+ as the reference group.
[27] We use the so-called *patent priority filing date*, which is the first time the applicant filed for a patent at *any* patent office.

quickly its use spread throughout the organisation. Therefore, we define our broadband variable as the number of years since activation to allow for timing flexibility.

We describe our headline regression in Equations 1-3, where subscript n and y refer to nodes and years respectively. We limit our sample to nodes that patented at least once during our 2000-2007 timeframe in order to minimise noise.[28]

$$pc_{ny} = \beta * BB_{n,y-1} + \phi_y + \epsilon_{ny} \qquad \text{Year FE} \tag{1}$$

$$pc_{ny} = \beta * BB_{n,y-1} + \phi_y + \theta_{nuts3} + \epsilon_{ny} \qquad \text{Region \& Year FE (YnR FE)} \tag{2}$$

$$BB_{n,y-1} = \gamma * RAIN_{n,y-2} + \phi_y + \theta_{nuts3} + \eta_{ny} \qquad \text{Rain IV} \tag{3}$$

pc_{ny} is the log patent count[29] and $BB_{n,y-1}$ is the number of years since broadband was activated, lagged one year and set to zero before activation. ϕ_y is a set of year dummies, present in all regressions, ϵ_{ny} is the error term. We will refer to the sparse specification in Equation 1 as the Year FE model.

In Equation 2 we add a set of NUTS3 fixed effects, θ_{nuts3}. We will refer to this specification as YnR FE, for year and region fixed effects. Equations 2 and 3 together define our rainfall instrumental variables approach, where $RAIN_{n,y-2}$ is the annual rainfall in a particular node n, lagged one year more for than the broadband variable as we assume investment decisions were based on past rainfall data.

We present results in Table 3, where we find positive coefficients for both fixed effects models (+0.07, columns 1 and 2), and an IV estimate that is considerably larger, at +0.17 (se 0.04). The IV coefficient implies that nodes with broadband access become 18.5% more prolific every year. Extrapolated to five years of access would mean an increase of $5 * 18.5 = 92.5\%$. Ten years: 185%. Clearly it is not reasonable to extrapolate this linear estimate until infinity.[30] Instead, we limit our extrapolation to observed lead times in broadband access as our coefficient is de facto identified through differences in broadband

[28] Note that the first LEs were activated in 2000 and by 2005 broadband diffusion exceeded 80% (see Figure 1).
[29] We add one to each count to retain observations with zero patenting that year.
[30] Unfortunately we can only identify one variable with one instrument, making it impossible to include e.g. a squared broadband access variable.

Table 3. Yearly patents per node (log), rainfall sample

	(1) Year FE	(2) YnR FE	(3) Rain IV
Years Broadband$_{t-1}$	0.07***	0.07***	0.17***
	(0.01)	(0.01)	(0.04)
Constant	0.11***	0.11***	
	(0.01)	(0.01)	
N	17,152	17,152	17,152
Year FE	✓	✓	✓
Region FE	×	✓	✓
First Stage F-Stat.			71.77

Notes: Standard errors clustered at the LE-level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

activation.

For each node n, we calculate the share of other nodes in the NUTS3 region that were activated g years later.[31] Taking averages over all nodes, we find that in 68% the comparison node activated in the same year or earlier. Of the remaining third that actually identifies the coefficient estimate, we find that almost 80% represents three or fewer years lead time. As a result, we do not feel confident to extrapolate the linear estimate more than two to three years, which corresponds to a 37-55% increase in patenting activity due to broadband access. Although this number appears large, it is not out of line with the literature. For example, Hegde, Herkenhoff and Zhu (2020) find that simply revealing patent information twelve months earlier already leads to an increase in patenting activity of 21%. Similarly, Bertschek, Cerquera and Klein (2013) estimate an increase of 56-63% in the probability of innovating due to broadband internet use in German firms. A potential explanation is that baseline patenting rates are very small - our predictions correspond to an average increase of just 0.6-0.9 patents per node per year.

As a robustness check, we turn to the demographic instruments Z_n in Equation 7. While the instruments are aggregated at the node level, we add the same set of characteristics

[31]We add nodes that activated earlier than the focal node in the zero gap category to avoid double counting.

aggregated at the MSOA level as control variables X_m, where subscript m refers to an MSOA. As a result, our sample goes from being node-year level to node-MSOA-year level.

$$pc_{mny} = \beta * BB_{n,y-1} + \phi_y + \epsilon_{mny} \qquad \text{Year FE} \tag{4}$$

$$pc_{mny} = \beta * BB_{n,y-1} + \phi_y + \theta_{nuts3} + \epsilon_{mny} \qquad \text{YnR FE} \tag{5}$$

$$pc_{mny} = \beta * BB_{n,y-1} + X_m\Lambda_2 + \phi_{y,2} + \theta_{nuts3,2} + \epsilon_{mny} \qquad \text{Demo IV } 2^{nd} \tag{6}$$

$$BB_{m,n,y-1} = Z_n\Gamma + X_m\Lambda_1 + \phi_{y,1} + \theta_{nuts3,1} + \eta_{mny} \qquad \text{Demo IV } 1^{st} \tag{7}$$

The fixed effects regressions of Equations 4 and 5 are identical to their rainfall sample equivalent, with the difference that the patent count pc_{mny} is now calculated at the intersection of nodes n and MSOA m. The instrumental variables setup now includes the node-level demographic instruments Z_n in the first stage and the MSOA-level demographic controls in both first and second stage (Equations 6 and 7).

Table 4. Yearly patents per node-msoa (log), demographics sample

	(1) Year FE	(2) YnR FE	(3) Demo IV
Years Broadband$_{t-1}$	0.02***	0.02***	0.05***
	(0.00)	(0.00)	(0.01)
N	31,168	31,168	31,168
Year FE	✓	✓	✓
Region FE	×	✓	✓
First Stage F-Stat.			338.89

Notes: Standard errors clustered at the LE-level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4 shows that while the estimates remain qualitatively similar, they are about a factor three smaller with a coefficient of +0.05 (se 0.01, column 3) for the demographic IV setup. Following the same logic as above, this implies an increase in yearly patenting of 10-15%. We are more confident in the exogeneity of the rainfall instrument and thus expect the true effect to be closer the rainfall estimate than the demographics ones, however it is reassuring that both lead to positive and significant results.

Table 5. Yearly citation-weighted patents per node (log). Rainfall and demographics samples

	(1) Year FE	(2) YnR FE	(3) Rain IV	(4) Demo IV
Years Broadband$_{t-1}$	0.17***	0.16***	0.50***	0.14***
	(0.01)	(0.02)	(0.10)	(0.02)
N	17 152	17 152	17 152	31 168
Year FE	✓	✓	✓	✓
Region FE	×	✓	✓	✓
MSOA controls	×	×	×	✓
First Stage F-Stat.			71.77	338.89

Notes: Standard errors clustered at the LE-level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Next, we weight patents by the number of forward citations, which are commonly used as a proxy for patent value. Table 5 shows that this increases the coefficients roughly threefold to +0.5 (se 0.1, Rain IV) and +0.14 (se 0.02, Demo IV). These coefficients combine the effect of broadband access on the number of yearly patents and the number of citations those patents receive. This suggests that not only does broadband lead to more patenting activity, it also leads to patents that are cited more widely, implying these are more relevant than patents generated without broadband access.

As a second robustness check we examine whether broadband access also increased the number of public patents (Table 6). These are patents generated by universities, government research labs and hospitals, which were already pre-broadband integrated into alternative international information system networks of academic institutions (BitNet, ARPAnet) and should thus be not, or less, affected by reductions in information access costs stemming from improvements in private network technology. We find estimates very close to zero for all specifications, with estimates that are significantly positive only in the pure fixed effects regression.

Table 6. Yearly *public* patents per node (log). Rainfall and demographics samples

	(1) Year FE	(2) YnR FE	(3) Rain IV	(4) Demo IV
Years Broadband$_{t-1}$	0.04***	0.02***	0.02	-0.00
	(0.00)	(0.00)	(0.03)	(0.01)
N	17 152	17 152	17 152	31 168
Year FE	✓	✓	✓	✓
Region FE	×	✓	✓	✓
MSOA controls	×	×	×	✓
First Stage F-Stat.			71.77	335.4

Notes: Standard errors clustered at the LE-level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

In the Appendix we show that our positive impact of broadband on patent counts and citations also holds when we restrict the sample to nodes with active patenting before the introduction of broadband (Tables 12 and 14) or analogously to patents by firms with patenting activity pre-broadband (Table 13 and 15).

C. *Historical placebo and concerns about instrument exogeneity*

In order to assess validity and investigate empirical support for the exogeneity of our instruments, we conduct a series of historical placebo tests on our data. If inventive activity in a local exchange area is perfectly orthogonal to broadband activation within the bounds of our exclusion restriction, our IV-approach should not predict significant differences for patenting across LE nodes in the absence of ADSL technology. For confirm this with a placebo test, we exploit the long trail of patent data available, going back to several decades before the arrival of broadband internet, and regress lagged, historical patenting in LEs on instrumented (future) broadband activation during the years 2000-2007.

Specifically, we use lagged patent applications filed ten and twenty years prior to the actual observation period. The results of this placebo test are reported in Table 7. Instrumenting earlier broadband activation within NUTS3-region based on variation in weather

Table 7. Historical placebo test for yearly patents per node (log). Rainfall and demographics samples

	10y-lagged # patents		*20y-lagged # patents*	
	(1) Rain IV	(2) Demo IV	(3) Rain IV	(4) Demo IV
Years Broadband$_{t-1}$	0.10***	0.06***	0.43***	0.04***
	(0.03)	(0.01)	(0.11)	(0.01)
N	17 152	31 168	17 152	31 168
Year FE	✓	✓	✓	✓
Region FE	✓	✓	✓	✓
MSOA controls	×	✓	×	✓
First Stage F-Stat.	71.77	113.69	71.77	113.69

Notes: Standard errors clustered at the LE-level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

conditions yields predictions of higher patenting activity in LEs already a decade before the actual arrival of ADSL at +0.1 (se 0.03, Rain IV) and even larger differences when looking at patenting 20-years prior to activitation (point estimate +0.4, se 0.1, Rain IV (3)). Also in models exploiting mismatch in demographic pull factors attracting ISPs between LEs and overlapping MSOAs as exclusion restriction, significantly higher rates of historical patenting are estimated for activated LEs (+0.06, se 0.01 for 10-year lagged patents and +0.04, se 0.01, for 20-year lags, Demo IV). The failure of this placebo test raises concerns regarding the exclusion restriction behind our identification strategy and warrants caution for the causal interpretation of our findings.

We consider possible explanations for this problem. One issue might be the strong concentration of patenting activity in relatively densely populated areas, which are on average less rainy. While this relationship is not necessarily causal, it might generate spurious correlations between our instruments and patenting in parts of our finite sample. This does not strictly need to imply a violation of exogeneity, however it is likely to decrease the power of instruments particularly in those areas were there is the highest variation in patenting. Our second concern relates to difficulties of our econometric IV specification to instrument the timing of treatment in our nodel-level panel: While our endogenous measure of years since broadband activation implies a time-varying intensity of treatment, the prediction out the first stage in our 2SLS model is de-facto cross-sectional, i.e. has the same value within a node in pre- and post-treatment stages, given our cross-sectional instruments. Using time-varying covariates in the first stage, on the other hand, would make the source of identification in our model unclear. Related to this, it might be problematic with regards to our linear specification that the treatment of broadband arrival is binary but irreversible: Being an attractive state, no LE reverts back to ISDN status after ADSL is introduced. Hence, in our panel the treatment is always 0 until it becomes 1, and always 1 afterwards within each node. Simple linear regression, as we use it in our first stage, might not be well suited for such type of data.

We attempt to solve these concerns in several ways. First, we devise an alternative version of our simple rainfall instrument, based on the interaction of LE area average precipitation with the share of overall broadband activated LEs in the economy in a given year. The idea behind this approach is to take into account yearly investment decisions of ISPs in broadband expansion in the overall U.K. market, which should be uncorrelated to the characteristics of a specific LE area micro-market, but affect LE entry differently based on attractiveness. This individual LE specific likelihood further varies from period to period as a function of national allocated budget and the already activated nodes. Second, instead of predicting *years since* broadband with 2SLS, we conduct a stepwise regression approach in which we, first, predict the actual year of activation based on our instruments and use this prediction to compute time since (predicted) activation in each year of observation for each node, and manually correct standard errors to account for this two-stage procedure in the second stage regression. Finally, in order to accommodate the specific nature of the treatment our data in a more suitable econometric specification, we devise a control function approach which uses a non-linear first stage based on a duration model, where the failure state is represented by the activation of broadband technology.[32]

[32] In the literature, such control function approaches are often found with reversed stages, e.g. where a linear first stage predicts a certain treatment and a non-linear second stage is used to estimate survival of a patient conditional

Specifically, we estimate a Cox proportional hazard model using our exclusion restrictions to estimate ADSL failure duration for each LE, and use the computed inverse Mills ratios from this stage to derive a correction term for the endogeneity affecting years since broadband activation in the second stage.

However, none of the above attempts is able to solve the problematic pre-trends detected through the historical placebo tests on our instrumental variable approach.[33] While similar types of instruments are widely established in the literature, this persisting problem constitutes an unsolved limitation regardig the causal interpretability of IV estimates, to be considered also in the remainder of the current version of this paper.

D. Mechanisms

As main mechanism we put forward that broadband access facilitates the inventive search process, in particular the acquisition of foreign knowledge becomes much easier with high speed internet as it allows inventors to browse through the body of international patents and technical news reports, as described in section I.B. In this section we show that patents originating from broadband activated nodes cite significantly more geographically distant prior art. Moreover, the international patent literature they cite is considerably more recent than that cited by non-broadband patents.

We start by constructing a dataset linking focal citing patents to applicant-added cited patents.[34] We limit our set of focal patents to those granted by the USPTO to avoid issues arising from differences in citation practices between offices. We consider the full universe of globally granted patents for the cited side to avoid selecting only highly internationalised patents. For each focal and cited patent, we obtain address information from the USPTO,

on receiving the treatment. In our case, however, the duration model appears more suitable to estimate the first stage, given the time delays in activation across nodes and the resulting staggered treatments.

[33] All estimation results from these attempts unreported in the paper.

[34] In line with Alcacer and Gittelman (2006), we only consider applicant-added citations to reflect knowledge flows. We will use examiner-added citations as a placebo test. We are able to distinguish between applicant and examiner added citations for 99.31% of the patents in our sample period. We exclude the remaining 0.7%. We also exclude inventor and firm self citations based on PSN inventors names in PATSTAT and BvD Orbis firm identifiers, as those reflect other types of knowledge flows than the ones we're interested in (similar to, e.g., Thompson (2006); Singh and Marx (2013)).

EP and IPO, retaining the most precise location.[35]

From this dataset, we generate our dependent variable $match_{ij}^{country}$, which equals one when the focal patent i and cited patent j are both from the same country. As discussed in chapter II, the location of a patent is an ambiguous construct given that patent inventors can reside in multiple places. To solve this, we assign patents to countries by randomly selecting one listed inventor and using their location.[36] Our results are robust to using a more conservative spillover measure that is based on the closest distance between locations across all possible combinations of cited-citing inventor pairs.

(8)
$$match_{ij}^{country} = \beta * BB_{i,y-1} + \phi_y + \psi_c + \epsilon_{ij} \qquad \text{Class and Year FE (CY FE)}$$

(9)
$$match_{ij}^{country} = \beta * BB_{i,y-1} + \phi_y + \psi_c + \theta_{nuts3} + \epsilon_{ij} \qquad \text{CY and Region FE (CYR FE)}$$

(10)
$$BB_{i,y-1} = \gamma * RAIN_{i,y-2} + \phi_y + \psi_c + \theta_{nuts3} + \eta_{ij} \qquad \text{Rain IV, } 1^{st}$$

(11)
$$match_{ij}^{country} = \beta * BB_{n,y-1} + X_{i_m}\Lambda_2 + \phi_{y,2} + \psi_{c,2} + \theta_{nuts3,2} + \epsilon_{ij} \qquad \text{Demo IV } 2^{nd}$$

(12)
$$BB_{m,n,y-1} = Z_{i_n}\Gamma + X_{i_m}\Lambda_1 + \phi_{y,1} + \psi_{c,1} + \theta_{nuts3,1} + \eta_{ij} \qquad \text{Demo IV } 1^{st}$$

Equations 8-12 describe our regression equations. We add primary 3-digit USPC technology class fixed effects ψ_c to absorb common differences in citing behaviour between technological domains. The rainfall and demographics IV now share the same sample, as both the node n and the MSOA m are fully determined by the location of the focal patent

[35]For example, many UK-based patents provide only city information in their USPTO document, whereas the IPO patent has street-level information. For non-UK patents this is less important, as there we require only country information.

[36]This approach is established practice in the literature starting with Jaffe, Trajtenberg and Henderson (1993). We follow Thompson (2006); Singh and Marx (2013) in randomly selecting one inventor location in case a patent lists multiple.

i.

Table 8 shows the results. We weight each observation such that all focal patents get a combined weight of one, regardless of the number of patent they cite.[37] With coefficients of -0.05 (se 0.03) and -0.02 (se 0.01) in the rainfall and demographics IV respectively, we find that access to broadband reduces the probability of citing domestic rather than foreign patents by 4-14%.[38]

Table 8. Probability of citing a same country patent

	(1) CY FE	(2) CYR FE	(3) Rain IV	(4) Demo IV
Years Broadband$_{t-1}$	-0.01***	-0.01	-0.05*	-0.02*
	(0.00)	(0.00)	(0.03)	(0.01)
N	63,633	63,633	63,633	63,633
Year + Class FE	✓	✓	✓	✓
Region FE	×	✓	✓	✓
MSOA controls	×	×	×	✓
First Stage F-Stat.			18.31	31.16

Notes: Standard errors clustered at the LE-level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The international prior art that they cite also appears to be more recent as we show in Table 9. The dependent variable is now $citage_{ij}$, the number of years between the application date of the focal patent i and each of its cited patents j. The coefficient of the rainfall IV in column (3) suggests that inventors with broadband access cite foreign patents which are 2-3 years more recent. This would explain why we find such a large increase in the number of yearly patents, given that Hegde, Herkenhoff and Zhu (2020) find a 21% increase in patenting due to a one-year reduction in the citation lag and our reduction is 2-3 times larger.[39]

[37] This avoids giving excess weight to focal patents that cite more often than others.

[38] The lower bound is the Demo IV estimate times two years, the upper bound is the Rain IV estimate times three years. See Section III.B for a discussion on why we do not extrapolate the results beyond 2-3 years.

[39] Note also that roughly 88% of all cited patents of UK focal patents are foreign, suggesting that most knowledge is sourced abroad.

Table 9. Citation lag in years, foreign cited patents only

	(1) Class+Year FE	(2) Region FE	(3) Rain IV	(4) Demo IV
Years Broadband$_{t-1}$	-0.12**	-0.11*	-0.93**	-0.05
	(0.05)	(0.06)	(0.47)	(0.16)
N	52,788	52,788	52,788	52,788
Year + Class FE	✓	✓	✓	✓
Region FE	×	✓	✓	✓
MSOA controls	×	×	×	✓
First Stage F-Stat.			20.57	30.14

Notes: Standard errors clustered at the LE-level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

We turn to examiner added citations to provide further evidence that these effects are truly causal. As the examiners are located in the art units of the USPTO, their behaviour should not be affected in any way by the broadband access of the inventor. This is exactly what we find in Tables 10 and 11.[40] We find insignificant results across the board, both for rainfall and the demographics IVs. The null effect on geographical spillovers is very precisely estimated. Unfortunately the citation lag effects have larger standard errors. It is nevertheless reassuring that the coefficient from the rainfall IV is almost cut in half for the examiner-added patents.

In the Appendix, we show that there is no corresponding reduction in the citation lag to same country patents (Table 16), suggesting that local knowledge already flowed relatively smoothly between inventors. We also do not find an increase in the total number of references (Table 17), which rules out inventors simply cite more international patents for legal reasons. Instead, this finding is fully consistent with our hypothesis of increased patenting activity through quicker access to foreign knowledge.

Our results hold when we restrict the sample to firms already patenting before broadband (Table 18, examiner robustness check in Table 19), ruling out that our results are driven by firms moving to internet activated regions. The geographic spillover effects are strongest

[40] We include art unit and USPTO filing year fixed effects in all specifications referring to examiner-added citations.

TABLE 10. PROBABILITY OF *examiner* ADDING A SAME COUNTRY PATENT

	(1) CY FE	(2) CYR FE	(3) Rain IV	(4) Demo IV
Years Broadband$_{t-1}$	-0.00	-0.00	-0.02	0.01
	(0.00)	(0.00)	(0.02)	(0.01)
N	34,155	34,155	34,155	34,155
Year + Class FE	✓	✓	✓	✓
Region FE	×	✓	✓	✓
MSOA controls	×	×	×	✓
First Stage F-Stat.			21.80	38.65

Notes: Standard errors clustered at the LE-level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE 11. CITATION LAG OF *examiner added citations* IN YEARS, FOREIGN CITED PATENTS ONLY

	(1) CY FE	(2) CYR FE	(3) Rain IV	(4) Demo IV
Years Broadband$_{t-1}$	-0.05	-0.08	-0.50	-0.21
	(0.06)	(0.06)	(0.51)	(0.21)
N	29,534	29,534	29,534	29,534
Year + Class FE	✓	✓	✓	✓
Region FE	×	✓	✓	✓
MSOA controls	×	×	×	✓
First Stage F-Stat.			20.00	35.98

Notes: Standard errors clustered at the LE-level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

in small firms, although the estimates are imprecise and only significant at 5% when we use the demographic instruments (Table 20).

We do not find any impact on the propensity to cite science (Table 21, only pre broadband firms in Table 22). This is reassuring, as scientific knowledge remained locked behind paywalls forming considerable barriers to entry.

We also do not find an increase in the number of inventors per patent, ruling out that the increase in patenting activity is driven by easier collaboration (Tables 23 and 24). This is in line with expectations, as the real increase in communication options arrived only with the second and third generation of broadband, while we only study the first generation. In the same vein, we do not find any effect on the number of international inventor teams (Tables 28 and 29).

We do find weak evidence that internet diffusion led to an increase in the number of new inventors, although the coefficient is only significant at 5% for the rainfall IV (Tables 25 and 26). This effect appears to be driven by small firms (Table 27). This would suggest broadband reduced the entry barriers for innovation - where before inventors needed a large personal network to get access to the relevant information, now the internet allows newcomers to find that information online more rapidly.

In Tables 32 - 43 in Appendix, we document that increases in citations to foreign prior art were particularly strong for references to EP patents. This suggests that the additional knowledge that inventors acquired from foreign patents due to broadband access was extracted more than expected from European Patent Office documents, and less from USPTO patents. This is particularly interesting, given the historically higher reliance of U.K. inventors on the USPTO patent body, as stressed by prior literature.

In additional, unreported, estimations we find no evidence for a significant reduction of inventor or firm self-citations in prior art references. We further cannot find evidence that broadband activation would have benefited access from firm inventors to university or public research organization patents. While effective knowledge transfer from academic research to industry is a particular concern to public R&D policy, prior literature has pointed out knowledge in patents from academic institutions to be particularly difficult to absorb for firms, which would be in line with our observation. Unlike expected, we also do not find any evidence that inventors would explore new knowledge outside of their domains or would become more likely to build on more distant fields after broadband access. This is in contrast to findings of related studies by Furman and Stern (2011); Furman, Nagler and Watzinger (2018); Biasi and Moser (2018); Iaria, Schwarz and Waldinger (2018);

Hegde, Herkenhoff and Zhu (2020). One attempt for explanation of this pattern could be that ADSL access might have become more inadvertently integral part of inventor search routines, while incurring the cost of visiting a library, or a biological resource centre, would usually precondition a specific problem requiring targeted search into a specific, perhaps more distant or unfamiliar domain.

IV. Conclusion

Inventors are standing on the shoulders of giants. Understanding the importance of access to prior knowledge, and in particular the cost of accessing this information, is a major quest of the empirical literature on cumulative innovation.

In this paper, we investigate the role of information technologies in facilitating the assimilation of distant prior art, and their repercussions on inventive activity. We exploit idiosyncrasies of the roll-out of high-speed internet in the UK to showcase the impact of early ADSL diffusion on inventive search and local R&D productivity in 2,029 patent generating catchment areas of local exchange nodes of the British telecommunications distribution network between 2000 and 2007. Using complementary identification strategies based on push-pull factors to ISP market entry, orthogonal to inventive firm or inventor establishment, we find that access to broadband services increased yearly patenting by up to 50% in the first 2-3 years after broadband arrival, alongside with a similar gain future technological impact. We can ascribe this increase in patenting largely to easier access to geographically distant knowledge through broadband, as evidenced by an increase in the share of applicant added foreign citations, and a steep decrease in average citation lags to prior art patents from abroad. We further show that these changes are not reflected in public sector patenting and examiner-added citations, and exclude a number of competing explanations underlying the extensive margin effect.

Our findings highlight a strong influence of broadband technology on the efficiency of inventive search and the accrual of cross-border knowledge spillovers. This is consistent with the view that information access plays and important role for subsequent invention, and shows that the elasticity of the cost of access to external knowledge is large and substantial, even in the middle of the information age, where most relevant codified knowledge was online available.

Limitations of our study are mainly imposed - at the current stage - by the historical pre-trends in patenting activities detected wihtin our instrumental variable apporach, which warrant caution for the causal interpretability of our results, and require further assessment. Other limitations relate to the comparatively short sample window, as the U.K. early broadband roll-out was essentially completed within a few years. Therefore, we are bound to extrapolating from a linear effect over this period, but our research setting does not allow to identify non-linearity in the impact of broadband internet over a longer time frame. Furthermore, our instrumental variable approach unfortunately lacks the power to investigate heterogeneity of results in greater detail. Alternative settings and different identification will be needed to explore the multifacetedness of main effects further. Finally, the generalizability of our findings is confined by the composition of our sample, which is tributary to the accuracy of patent address information and limited to inventive activity geographically concentrated in the UK.

REFERENCES

Aghion, Philippe, and Peter Howitt. 1992. “A model of growth through creative destruction.” *Econometrica*, 60(2): 323–351.

Agrawal, Ajay, and Avi Goldfarb. 2008. “Restructuring research: Communication costs and the democratization of university innovation.” *American Economic Review*, 98(4): 1578–90.

Ahlfeldt, Gabriel, Pantelis Koutroumpis, and Tommaso Valletti. 2017. “Speed 2.0: Evaluating access to universal digital highways.” *Journal of the European Economic Association*, 15(3): 586–625.

Akcigit, Ufuk, Salome Baslandze, and Stefanie Stantcheva. 2016. “Taxation and the international mobility of inventors.” *American Economic Review*, 106(10): 2930–81.

Alcacer, Juan, and Michelle Gittelman. 2006. “Patent citations as a measure of knowledge flows: The influence of examiner citations.” *The Review of Economics and Statistics*, 88(4): 774–779.

Baruffaldi, Stefano, and Markus Simeth. 2020. “Patents and Knowledge Diffusion: The Effect of Early Disclosure.” *Research Policy*, forthcoming.

Bertschek, Irene, Daniel Cerquera, and Gordon J Klein. 2013. “More bits–more bucks? Measuring the impact of broadband internet on firm performance.” *Information Economics and Policy*, 25(3): 190–203.

Biasi, Barbara, and Petra Moser. 2018. “Effects of Copyrights on Science - Evidence from the US Book Republication Program.” National Bureau of Economic Research Working Paper 24255.

Black, S E, and L M Lynch. 2001. “How to compete: The impact of workplace practices and information technology on productivity.” *Review of Economics and Statistics*, 83(3): 434–445.

Bloom, Nicholas, Mark Schankerman, and John Van Reenen. 2013. "Identifying technology spillovers and product market rivalry." *Econometrica*, 81(4): 1347–1393.

Bouckaert, Jan, Theon van Dijk, and Frank Verboven. 2010. "Access regulation, competition, and broadband penetration: An international study." *Telecommunications Policy*, 34(11): 661–671.

Bryan, Kevin A, and Yasin Ozcan. 2016. "The impact of open access mandates on invention." *Mimeo, Toronto.*

Cheyre, Cristobal, Steven Klepper, and Francisco Veloso. 2015. "Spinoffs and the mobility of US merchant semiconductor inventors." *Management Science*, 61(3): 487–506.

Cohen, Wesley M, Richard R Nelson, and John P Walsh. 2000. "Protecting Their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (or Not)." National Bureau of Economic Research Working Paper 7552.

Czernich, Nina, Oliver Falck, Tobias Kretschmer, and Ludger Woessmann. 2011. "Broadband infrastructure and economic growth." *The Economic Journal*, 121(552): 505–532.

de Rassenfosse, G., G. Pellegrino, and E. Raiteri. 2020. "Do patents enable disclosure? Evidence from the 1951 Invention Secrecy Act." Ecole polytechnique fédérale de Lausanne.

de Rassenfosse, G., J. Kozak, and F Seliger. 2019. "Geocoding of worldwide patent data." *Nature Scientific Data*, 6(260).

Devlin, Alan. 2009. "The misunderstood function of disclosure in patent law." *Harvard Journal of Law & Technology*, 23: 401.

Dewan, S, and KL Kraemer. 2000. "Information technology and productivity: Evidence from country-level data." *Management Science*, 46(4): 548–562.

Dhyne, Emmanuel, Jozef Konings, Jeroen Van den bosch, and Stijn Vanormelingen. 2018. "The Return on Information Technology: Who Benefits Most?" CEPR Discussion Paper No. DP13246.

Ding, Waverly W, Sharon G Levin, Paula E Stephan, and Anne E Winkler. 2010. "The impact of information technology on academic scientists' productivity and collaboration patterns." *Management Science*, 56(9): 1439–1461.

Dittmar, Jeremiah E. 2011. "Information technology and economic change: the impact of the printing press." *The Quarterly Journal of Economics*, 126(3): 1133–1172.

Forman, Chris, and Nicolas van Zeebroeck. 2012. "From wires to partners: How the Internet has fostered R&D collaborations within firms." *Management science*, 58(8): 1549–1568.

Forman, Chris, and Nicolas van Zeebroeck. 2019. "Digital technology adoption and knowledge flows within firms: Can the Internet overcome geographic and technological distance?" *Research policy*, 48(8): 103697.

Furman, Jeffrey L., and Scott Stern. 2011. "Climbing atop the Shoulders of Giants: The Impact of Institutions on Cumulative Research." *American Economic Review*, 101(5): 1933–63.

Furman, Jeffrey L, Markus Nagler, and Martin Watzinger. 2018. "Disclosure and subsequent innovation: Evidence from the patent depository library program." National Bureau of Economic Research.

Gavazza, Alessandro, Mattia Nardotto, and Tommaso Valletti. 2019. "Internet and politics: Evidence from UK local elections and local government policies." *The Review of Economic Studies*, 86(5): 2092–2135.

Giuri, Paola, Myriam Mariani, Stefano Brusoni, Gustavo Crespi, Dominique Francoz, Alfonso Gambardella, Walter Garcia-Fontes, Aldo Geuna, Raul

Gonzales, Dietmar Harhoff, and Karin Hoisl. 2007. "Inventors and invention processes in Europe: Results from the PatVal-EU survey." *Research Policy*, 36(8): 1107–1127.

Graham, S, and D Hegde. 2015. "Disclosing patents' secrets." *Science*, 347(6219): 236–237.

Griffith, Rachel, Rupert Harrison, and John Van Reenen. 2006. "How special is the special relationship? Using the impact of US R&D spillovers on UK firms as a test of technology sourcing." *American Economic Review*, 96(5): 1859–1875.

Griliches, Zvi. 1990. *Patent statistics as economic indicators: A survey, part 2.* NBER.

Hegde, Deepak, and Hong Luo. 2018. "Patent publication and the market for ideas." *Management Science*, 64(2): 652–672.

Hegde, Deepak, Kyle Herkenhoff, and Chenqi Zhu. 2020. "Patent disclosure and innovation." *Available at SSRN 3158031.*

Hjort, Jonas, and Jonas Poulsen. 2019. "The arrival of fast internet and employment in Africa." *American Economic Review*, 109(3): 1032–79.

Iaria, Alessandro, Carlo Schwarz, and Fabian Waldinger. 2018. "Frontier Knowledge and Scientific Production: Evidence from the Collapse of International Science*." *The Quarterly Journal of Economics*, 133(2): 927–991.

Jaffe, Adam B. 1986. "Technological opportunity and spillovers of R&D: evidence from firms' patents, profits and market value." national bureau of economic research.

Jaffe, Adam B, Manuel Trajtenberg, and Michael S Fogarty. 2000. "Knowledge spillovers and patent citations: Evidence from a survey of inventors." *American Economic Review*, 90(2): 215–218.

Jaffe, Adam B, Manuel Trajtenberg, and Rebecca Henderson. 1993. "Geographic localization of knowledge spillovers as evidenced by patent citations." *The Quarterly Journal of Economics*, 108(3): 577–598.

Jones, Charles I. 1995. "R&D-based models of economic growth." *Journal of Political Economy*, 103(4): 759–784.

Jones, Eric. 2003. *The European miracle: environments, economies and geopolitics in the history of Europe and Asia.* Cambridge University Press.

Keeble, David, and Lilach Nachum. 2001. "Why Do Business Service Firms Cluster? Small Consultancies, Clustering and Decentralisation in London and Southern England." Centre for Business Research, University of Cambridge Working Papers wp194.

Lemley, Mark A. 2012. "The myth of the sole inventor." *Mich. L. Rev.*, 110(5): 709.

Lück, Sonja, Benjamin Balsmeier, Florian Seliger, and Lee Fleming. 2019. "Early disclosure law reduces duplication in the US and European patent systems." Northwestern Pritzker School of Law.

Machlup, Fritz, and Edith Penrose. 1950. "The patent controversy in the nineteenth century." *The Journal of Economic History*, 10(1): 1–29.

MacKie-Mason, Jeffrey K, and Hal R Varian. 1996. "Some economics of the Internet."

Mokyr, Joel. 2005. "Long-term economic growth and the history of technology." In *Handbook of economic growth.* Vol. 1, 1113–1180. Elsevier.

Morrison, Greg, Massimo Riccaboni, and Fabio Pammolli. 2017. "Disambiguation of patent inventors and assignees using high-resolution geolocation data." *Scientific data*, 4: 170064.

Moser, Petra, and Alessandra Voena. 2012. "Compulsory Licensing: Evidence from the Trading with the Enemy Act." *American Economic Review*, 102(1): 396–427.

Nardotto, Mattia, Tommaso Valletti, and Frank Verboven. 2015. "Unbundling the incumbent: Evidence from UK broadband." *Journal of the European Economic Association*, 13(2): 330–362.

Ofcom. 2014. "Fixed access market reviews: wholesale local access, wholesale fixed analogue exchange lines, ISDN2 and ISDN30 (Annexes)."

Ouelette, Lisa L. 2012. "Do patents disclose useful information?" *Harvard Journal of Law & Technology*, 25(2): 545–608.

Pavitt, Keith, and Pari Patel. 1998. "The international distribution and determinants of technological activities." *Oxford Review of Economic Policy*, 4(4): 35–55.

Romer, Paul M. 1990. "Endogenous technological change." *Journal of political Economy*, 98(5, Part 2): S71–S102.

Sampat, Bhaven N. 2010. "When do applicants search for prior art?" *The Journal of Law and Economics*, 53(2): 399–416.

Scotchmer, Suzanne. 1991. "Standing on the shoulders of giants: cumulative research and the patent law." *Journal of economic perspectives*, 5(1): 29–41.

Scotchmer, Suzanne, and Jerry Green. 1990. "Novelty and disclosure in patent law." *The RAND Journal of Economics*, 131–146.

Singh, Jasjit. 2005. "Collaborative networks as determinants of knowledge diffusion patterns." *Management science*, 51(5): 756–770.

Singh, Jasjit. 2008. "Distributed R&D, cross-regional knowledge integration and quality of innovative output." *Research Policy*, 37(1): 77–96.

Singh, Jasjit, and Matt Marx. 2013. "Geographic constraints on knowledge spillovers: Political borders vs. spatial proximity." *Management Science*, 59(9): 2056–2078.

Spear, Brian. 1996. "Patent information on the internet." *World Patent Information*, 18(1): 33–37.

Stiroh, Kevin J. 2002. "Information technology and the U.S. productivtiy revival: What do the industry data say?" *American Economic Review*, 92(5): 1559–1576.

Thompson, Peter. 2006. "Patent citations and the geography of knowledge spillovers: evidence from inventor-and examiner-added citations." *The Review of Economics and Statistics*, 88(2): 383–388.

Verspagen, B, and W Schoenmakers. 2004. "The spatial dimension of patenting by multinational firms in Europe." *Journal of Economic Geography*, 4(1): 23–42.

Weitzman, Martin L. 1998. "Recombinant growth." *The Quarterly Journal of Economics*, 113(2): 331–360.

Williams, Heidi L. 2013. "Intellectual Property Rights and Innovation: Evidence from the Human Genome." *Journal of Political Economy*, 121(1): 1–27.

Zheng, Yanfeng, and Qinyu Wang. 2020. "Shadow of The Great Firewall: The Impact of Google Blockade on Innovation in China." *Strategic Management Journal, Forthcoming.*

APPENDIX A Tables

TABLE 12. YEARLY PATENTS PER NODE (LOG), RAINFALL SAMPLE - PRE-BROADBAND ACTIVE LES

	(1) Class+Year FE	(2) Region FE	(3) Rain IV
Years Broadband$_{t-1}$	0.08***	0.09***	0.23***
	(0.01)	(0.01)	(0.07)
Constant	0.21***	0.19***	
	(0.02)	(0.02)	
N	8712	8712	8712
Year FE	✓	✓	✓
Region FE	×	✓	✓
First Stage F-Stat.			40.51

Notes: Standard errors clustered at the LE-level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 13. Yearly patents per node (log), rainfall sample - pre-Broadband established firms

	(1) Class+Year FE	(2) Region FE	(3) Rain IV
Years Broadband$_{t-1}$	0.04***	0.04***	0.11***
	(0.01)	(0.01)	(0.03)
Constant	0.07***	0.06***	
	(0.01)	(0.01)	
N	17,152	17,152	17,152
Year FE	✓	✓	✓
Region FE	×	✓	✓
First Stage F-Stat.			71.77

Notes: Standard errors clustered at the LE-level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 14. Yearly patents per node (log), rainfall sample - pre-Broadband active LEs

	(1) Class+Year FE	(2) Region FE	(3) Rain IV
Years Broadband$_{t-1}$	0.19***	0.21***	0.70***
	(0.02)	(0.03)	(0.19)
Constant	0.60***	0.57***	
	(0.05)	(0.05)	
N	8,712	8,712	8,712
Year FE	✓	✓	✓
Region FE	×	✓	✓
First Stage F-Stat.			40.51

Notes: Standard errors clustered at the LE-level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 15. Yearly citation weighted patents per node (log), rainfall sample - pre-Broadband established firms

	(1) Class+Year FE	(2) Region FE	(3) Rain IV
Years Broadband$_{t-1}$	0.09***	0.09***	0.33***
	(0.01)	(0.01)	(0.08)
Constant	0.20***	0.19***	
	(0.02)	(0.02)	
N	17,152	17,152	17,152
Year FE	✓	✓	✓
Region FE	×	✓	✓
First Stage F-Stat.			71.77

Notes: Standard errors clustered at the LE-level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 16. Citation lag in years, domestic cited patents only

	(1) Class+Year FE	(2) Region FE	(3) Rain IV	(4) Demogs IV
Years Broadband$_{t-1}$	0.01	0.01	0.10	0.02
	(0.01)	(0.01)	(0.08)	(0.03)
Constant	1.56***	1.56***		
	(0.02)	(0.02)		
N	6,402	6,399	6,397	6,397
Year + Class FE	✓	✓	✓	✓
Region FE	×	✓	✓	✓
First Stage F-Stat.			18.85	31.83

Notes: Standard errors clustered at the LE-level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 17. Applicant-added citations per patent (log)

	(1) Class+Year FE	(2) Region FE	(3) Rain IV	(4) Demogs IV
Years Broadband$_{t-1}$	0.01	0.01	-0.16	-0.09*
	(0.02)	(0.02)	(0.13)	(0.06)
Constant	1.96***	1.96***		
	(0.05)	(0.04)		
N	6,402	6,399	6,397	6,397
Year + Class FE	✓	✓	✓	✓
Region FE	×	✓	✓	✓
First Stage F-Stat.			18.85	31.83

Notes: Standard errors clustered at the LE-level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 18. Probability of citing a same country patent - pre-broadband established firms

	(1) Class+Year FE	(2) Region FE	(3) Rain IV	(4) Demogs IV
Years Broadband$_{t-1}$	-0.01***	-0.01***	-0.07**	-0.05***
	(0.00)	(0.00)	(0.03)	(0.02)
N	40,398	40,398	40,398	40,398
Year + Class FE	✓	✓	✓	✓
Region FE	×	✓	✓	✓
First Stage F-Stat.			15.39	14.85

Notes: Standard errors clustered at the LE-level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 19. Probability of examiner adding a same country patent reference - pre-broadband established firms

	(1) Class+Year FE	(2) Region FE	(3) Rain IV	(4) Demogs IV
Years Broadband$_{t-1}$	-0.00*	-0.00	-0.03	0.01
	(0.00)	(0.00)	(0.04)	(0.02)
N	20,760	20,760	20,760	20,760
Year + Class FE	✓	✓	✓	✓
Region FE	×	✓	✓	✓
First Stage F-Stat.			15.39	14.85

Notes: Standard errors clustered at the LE-level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 20. Firm size splits (above median) for same country citations

	(1) Rain IV Small	(2) Rain IV Large	(3) Demogs IV Small	(4) Demogs IV Large
Years Broadband$_{t-1}$	-0.07	0.01	-0.04**	0.01
	(0.05)	(0.03)	(0.02)	(0.01)
N	33,458	30,164	33,458	30,169
Year + Class FE	✓	✓	✓	✓
Region FE	✓	✓	✓	✓
First Stage F-Stat.	13.65	12.67	11.68	31.06

Notes: Standard errors clustered at the LE-level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 21. Probability of reliance on science [01]

	(1) Class+Year FE	(2) Region FE	(3) Rain IV	(4) Demogs IV
Years Broadband$_{t-1}$	0.00	0.01	-0.05	-0.03
	(0.01)	(0.01)	(0.04)	(0.02)
Constant	0.29***	0.28***		
	(0.01)	(0.02)		
N	6,402	6,399	6,397	6,397
Year + Class FE	✓	✓	✓	✓
Region FE	×	✓	✓	✓
First Stage F-Stat.			18.85	31.83

Notes: Standard errors clustered at the LE-level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 22. Probability of reliance on science [01] - Pre-broadbrand established firms

	(1) Class+Year FE	(2) Region FE	(3) Rain IV	(4) Demogs IV
Years Broadband$_{t-1}$	0.01	0.01	-0.03	-0.04
	(0.01)	(0.01)	(0.06)	(0.03)
Constant	0.30***	0.29***		
	(0.02)	(0.02)		
N	4,027	4,017	4,015	4,015
Year + Class FE	✓	✓	✓	✓
Region FE	×	✓	✓	✓
First Stage F-Stat.			18.85	31.83

Notes: Standard errors clustered at the LE-level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 23. Number of inventors (log) per patent

	(1) Class+Year FE	(2) Region FE	(3) Rain IV	(4) Demogs IV
Years Broadband$_{t-1}$	0.00	0.01	0.03	0.03
	(0.01)	(0.01)	(0.06)	(0.03)
Constant	0.62***	0.61***		
	(0.03)	(0.02)		
N	8,301	8,300	8,298	8,298
Year + Class FE	✓	✓	✓	✓
Region FE	×	✓	✓	✓
First Stage F-Stat.			23.83	31.89

Notes: Standard errors clustered at the LE-level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 24. Number of inventors (log) per patent - pre-Broadband established firms

	(1) Class+Year FE	(2) Region FE	(3) Rain IV	(4) Demogs IV
Years Broadband$_{t-1}$	-0.00	0.01	-0.03	0.03
	(0.01)	(0.02)	(0.08)	(0.04)
Constant	0.69***	0.67***		
	(0.03)	(0.03)		
N	5,116	5,108	5,106	5,106
Year + Class FE	✓	✓	✓	✓
Region FE	×	✓	✓	✓
First Stage F-Stat.			14.46	16.54

Notes: Standard errors clustered at the LE-level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 25. Probability of new inventor entry [01]

	(1) Class+Year FE	(2) Region FE	(3) Rain IV	(4) Demogs IV
Years Broadband$_{t-1}$	0.01*	0.00	0.16**	0.03
	(0.01)	(0.01)	(0.06)	(0.02)
Constant	0.64***	0.66***		
	(0.02)	(0.02)		
N	6402	6399	6397	6397
Year + Class FE	✓	✓	✓	✓
Region FE	×	✓	✓	✓
First Stage F-Stat.			18.85	31.83

Notes: Standard errors clustered at the LE-level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 26. Probability of new inventor entry [01] - pre-broadband established firms

	(1) Class+Year FE	(2) Region FE	(3) Rain IV	(4) Demogs IV
Years Broadband$_{t-1}$	0.02*	0.01	0.20**	0.01
	(0.01)	(0.01)	(0.08)	(0.04)
Constant	0.61***	0.62***		
	(0.02)	(0.03)		
N	4,027	4,017	4,015	4,015
Year + Class FE	✓	✓	✓	✓
Region FE	×	✓	✓	✓
First Stage F-Stat.			15.36	14.17

Notes: Standard errors clustered at the LE-level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 27. Firm size splits for new inventor entry [01]

	(1) Rain IV Small	(2) Rain IV Large	(3) Demogs IV Small	(4) Demogs IV Large
Years Broadband$_{t-1}$	0.28**	0.09	0.04	0.02
	(0.12)	(0.07)	(0.05)	(0.03)
N	3325	2976	3325	2976
Year + Class FE	✓	✓	✓	✓
Region FE	✓	✓	✓	✓
First Stage F-Stat.	11.84	13.92	11.78	30.54

Notes: Standard errors clustered at the LE-level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 28. Probability of geographic co-location of inventor teams (same country)

	(1) Class+Year FE	(2) Region FE	(3) Rain IV	(4) Demogs IV
Years Broadband$_{t-1}$	-0.01	-0.01	-0.02	-0.01
	(0.01)	(0.01)	(0.06)	(0.03)
Constant	0.25***	0.24***		
	(0.02)	(0.02)		
N	8,301	8,300	8,298	8,298
Year + Class FE	✓	✓	✓	✓
Region FE	×	✓	✓	✓
First Stage F-Stat.			23.83	31.89

Notes: Standard errors clustered at the LE-level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 29. Probability of geographic co-location of inventor teams (same country) - pre-broadband established firms

	(1) Class+Year FE	(2) Region FE	(3) Rain IV	(4) Demogs IV
Years Broadband$_{t-1}$	-0.02*	-0.01	-0.05	-0.03
	(0.01)	(0.01)	(0.08)	(0.04)
Constant	0.25***	0.24***		
	(0.03)	(0.02)		
N	5,116	5,108	5,106	5,106
Year + Class FE	✓	✓	✓	✓
Region FE	×	✓	✓	✓
First Stage F-Stat.			14.46	16.54

Notes: Standard errors clustered at the LE-level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 30. Probability of new inventor collaboration

	(1) Class+Year FE	(2) Region FE	(3) Rain IV	(4) Demogs IV
Years Broadband$_{t-1}$	0.00	0.01	0.06	0.06***
	(0.01)	(0.01)	(0.05)	(0.02)
Constant	0.50***	0.49***		
	(0.02)	(0.02)		
N	8,301	8,300	8,298	8,298
Year + Class FE	✓	✓	✓	✓
Region FE	×	✓	✓	✓
First Stage F-Stat.			23.83	31.89

Notes: Standard errors clustered at the LE-level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 31. Probability of new inventor collaboration - pre-broadband established firms

	(1) Class+Year FE	(2) Region FE	(3) Rain IV	(4) Demogs IV
Years Broadband$_{t-1}$	0.00	0.01	-0.00	0.07**
	(0.01)	(0.01)	(0.07)	(0.03)
Constant	0.54***	0.52***		
	(0.02)	(0.03)		
N	5,116	5,108	5,106	5,106
Year + Class FE	✓	✓	✓	✓
Region FE	×	✓	✓	✓
First Stage F-Stat.			14.46	16.54

Notes: Standard errors clustered at the LE-level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 32. Probability of applicant citing a US patent document

	(1) Class+Year FE	(2) Region FE	(3) Rain IV	(4) Demo IV
Years Broadband$_{t-1}$	0.00	-0.00	-0.07*	-0.00
	(0.00)	(0.00)	(0.04)	(0.01)
N	63,638	63,638	63,638	63,638
Year + Class FE	✓	✓	✓	✓
Region FE	×	✓	✓	✓
MSOA controls	×	×	×	✓
First Stage F-Stat.			18.31	31.16

Notes: Standard errors clustered at the LE-level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 33. Probability of applicant citing a *foreign* US patent document

	(1) Class+Year FE	(2) Region FE	(3) Rain IV	(4) Demo IV
Years Broadband$_{t-1}$	-0.00	-0.00	-0.10**	-0.02
	(0.00)	(0.00)	(0.04)	(0.01)
N	56,028	56,028	56,028	56,028
Year + Class FE	✓	✓	✓	✓
Region FE	×	✓	✓	✓
MSOA controls	×	×	×	✓
First Stage F-Stat.			18.77	30.26

Notes: Standard errors clustered at the LE-level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 34. Probability of *examiner added* citation to US patent document

	(1) Class+Year FE	(2) Region FE	(3) Rain IV	(4) Demo IV
Years Broadband$_{t-1}$	-0.00	-0.00	0.01	-0.00
	(0.00)	(0.00)	(0.01)	(0.01)
N	34,161	34,161	34,161	34,161
Year + Class FE	✓	✓	✓	✓
Region FE	×	✓	✓	✓
MSOA controls	×	×	×	✓
First Stage F-Stat.			21.80	38.65

Notes: Standard errors clustered at the LE-level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 35. Probability of *examiner added* citation to *foreign* US patent document

	(1) Class+Year FE	(2) Region FE	(3) Rain IV	(4) Demo IV
Years Broadband$_{t-1}$	-0.00	-0.00	-0.00	-0.00
	(0.00)	(0.00)	(0.01)	(0.01)
N	31,876	31,876	31,876	31,876
Year + Class FE	✓	✓	✓	✓
Region FE	×	✓	✓	✓
MSOA controls	×	×	×	✓
First Stage F-Stat.			19.31	38.67

Notes: Standard errors clustered at the LE-level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 36. Probability of applicant citing a US patent document - pre-broadband established firms only

	(1) Class+Year FE	(2) Region FE	(3) Rain IV	(4) Demo IV
Years Broadband$_{t-1}$	0.00	-0.00	-0.03	-0.00
	(0.01)	(0.01)	(0.04)	(0.02)
N	40,402	40,402	40,402	40,402
Year + Class FE	✓	✓	✓	✓
Region FE	×	✓	✓	✓
MSOA controls	×	×	×	✓
First Stage F-Stat.			15.39	14.85

Notes: Standard errors clustered at the LE-level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE 37. PROBABILITY OF APPLICANT CITING A *foreign* US PATENT DOCUMENT - PRE-BROADBAND ESTABLISHED FIRMS ONLY

	(1) Class+Year FE	(2) Region FE	(3) Rain IV	(4) Demo IV
Years Broadband$_{t-1}$	-0.00	-0.01	-0.07	-0.03
	(0.01)	(0.01)	(0.04)	(0.02)
N	35,659	35,659	35,659	35,659
Year + Class FE	✓	✓	✓	✓
Region FE	×	✓	✓	✓
MSOA controls	×	×	×	✓
First Stage F-Stat.			15.86	14.60

Notes: Standard errors clustered at the LE-level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE 38. PROBABILITY OF APPLICANT CITING AN EP PATENT DOCUMENT

	(1) Class+Year FE	(2) Region FE	(3) Rain IV	(4) Demo IV
Years Broadband$_{t-1}$	0.00	0.00	0.04	0.01
	(0.00)	(0.00)	(0.02)	(0.01)
N	63,638	63,638	63,638	63,638
Year + Class FE	✓	✓	✓	✓
Region FE	×	✓	✓	✓
MSOA controls	×	×	×	✓
First Stage F-Stat.			18.31	31.16

Notes: Standard errors clustered at the LE-level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 39. Probability of applicant citing an *foreign* EP patent document

	(1) Class+Year FE	(2) Region FE	(3) Rain IV	(4) Demo IV
Years Broadband$_{t-1}$	-0.00	0.00	0.05**	0.01
	(0.00)	(0.00)	(0.03)	(0.01)
N	56,028	56,028	56,028	56,028
Year + Class FE	✓	✓	✓	✓
Region FE	×	✓	✓	✓
MSOA controls	×	×	×	✓
First Stage F-Stat.			18.77	30.26

Notes: Standard errors clustered at the LE-level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 40. Probability of *examiner added* citation to EP patent document

	(1) Class+Year FE	(2) Region FE	(3) Rain IV	(4) Demo IV
Years Broadband$_{t-1}$	-0.00	-0.00	0.01	-0.00
	(0.00)	(0.00)	(0.01)	(0.01)
N	34,161	34,161	34,161	34,161
Year + Class FE	✓	✓	✓	✓
Region FE	×	✓	✓	✓
MSOA controls	×	×	×	✓
First Stage F-Stat.			21.80	38.65

Notes: Standard errors clustered at the LE-level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 41. Probability of *examiner added* citation to *foreign* EP patent document

	(1) Class+Year FE	(2) Region FE	(3) Rain IV	(4) Demo IV
Years Broadband$_{t-1}$	-0.00	-0.00	-0.01	0.00**
	(0.00)	(0.00)	(0.01)	(0.00)
N	31,876	31,876	31,876	31,876
Year + Class FE	✓	✓	✓	✓
Region FE	×	✓	✓	✓
MSOA controls	×	×	×	✓
First Stage F-Stat.			19.31	38.67

Notes: Standard errors clustered at the LE-level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 42. Probability of applicant citing an EP patent document - pre-broadband established firms only

	(1) Class+Year FE	(2) Region FE	(3) Rain IV	(4) Demo IV
Years Broadband$_{t-1}$	0.00	0.00	0.05*	0.02
	(0.00)	(0.00)	(0.03)	(0.01)
N	40,402	40,402	40,402	40,402
Year + Class FE	✓	✓	✓	✓
Region FE	×	✓	✓	✓
MSOA controls	×	×	×	✓
First Stage F-Stat.			15.39	14.85

Notes: Standard errors clustered at the LE-level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 43. Probability of applicant citing a *foreign* EP patent document - pre-broadband established firms only

	(1) Class+Year FE	(2) Region FE	(3) Rain IV	(4) Demo IV
Years Broadband$_{t-1}$	-0.00	0.00	0.07*	0.03*
	(0.00)	(0.01)	(0.04)	(0.02)
N	35,659	35,659	35,659	35,659
Year + Class FE	✓	✓	✓	✓
Region FE	×	✓	✓	✓
MSOA controls	×	×	×	✓
First Stage F-Stat.			15.86	14.60

Notes: Standard errors clustered at the LE-level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

APPENDIX B Figures

FIGURE 5. NUTS3 REGIONS (COUNTIES) OF GREAT BRITAIN (N=167)

Figure 6. Census 2001 MSOA areas across Great Britain (N=3,374)