THE GEOGRAPHICAL EXTENT OF KNOWLEDGE SPILLOVERS IN THE UK, AS MEASURED THROUGH PATENT CITATIONS

Philip Wilkinson

09/08/20

MSc Smart Cities and Urban Analytics Dissertation – CASA0010

Word count: 11,764

Supervisor: Dr Elsa Arcaute

This dissertation is submitted in part requirement for the MSc in the Centre for Advanced Spatial Analysis, Bartlett Faculty of the Built Environment, UCL

Abstract

Agglomeration economies have been studied from a wide variety of different disciplines, emphasising different influences and factors that attract firms to locating near to each other, creating a vast corpus of literature on the subject. Most modern literature, however, acknowledges the central role that knowledge and knowledge spillovers have in modern agglomeration economies and hence how this may foster local, regional, and national economic growth. This paper therefore continues to expand on the existing literature by analysing patent citations to explore knowledge spillovers within the UK in terms of how distance and administrative boundaries influence the probability of citation. In doing so, this paper finds that local and regional institutional factors constrain knowledge spillovers locally, that most knowledge spillovers are exhausted within an extended commuting boundary, and that the effect of distance has increased over time. Consequently, it is recommended that policy solutions must acknowledge the role of local institutions in fostering knowledge spillovers, that communication networks between clusters of firms and temporary agglomerations may reduce the effects of distance, and that geography affects knowledge spillovers differently across industries. Future research could continue to expand on this by exploring European wide knowledge spillovers or by integrating different proximity measures into the analysis to further our understanding of knowledge spillovers in the UK and Europe.

Acknowledgements

I would like to thank Dr Elsa Arcaute and acknowledge her support and guidance throughout this project as my supervisor, including help with the overall direction and structure of the dissertation. I would also like acknowledge the help and advice of Professor Michael Barrett and Dr Karl Prince for direction, Dr Andrew Machlachlan for support, and Gabrielle Anderson for structural advice.

Declaration

I, Philip Wilkinson, hereby declare that this dissertation is all my own original work and that all sources have been acknowledged. It is 11,764 words in length as computed by Microsoft Word.

Signed:

Date: 09/08/2020

Contents

List of figures 5
List of tables 6
List of acronyms and abbreviations 7
1) Introduction 8
2) Literature review 10
3) Patent data exploration 20
3.1) Statement of ethics 27
4) Methodology 27
5) Results 30
5.1) Binomial t-test results 30
5.2) Baseline regression estimation results 31
5.3) The effects of technology on knowledge spillovers 35
5.4) The effects of time on knowledge spillovers 39
5.5) The effects of distance on knowledge spillovers in different industries 41
6) Implications and discussion 44
7) Conclusion 50
Bibliography 53
Appendices 60
Appendix A 60
Appendix B 62
Appendix C 63
Appendix D 67
Research log 70

List of figures

Figure 1: *The local (NUTS3) sum of patent inventor share for individuals from 1977 to 2019 whose entire inventor team is located in the UK. This refers to the share of the patent assigned to the individuals when the patent is applied for ………………………………………………………………………………… 21*

Figure 2: *The regional distribution of the sum of patent inventor share by individuals from 1977 to 2019 whose entire inventor team is based in the UK and whose patent has been cited by applications under the PCT or the EPO …………………………………………………………………………………………………….. 22*

Figure 3: The regional distribution of the sum of patent inventor share by individuals from 1977 to 2019 from patents that cite UK patents. The regions that produce the most citations are shown here with a) UK, b) Europe, c) United States, d) China, South Korea and Japan ……………………………. 23/24

Figure 4: A histogram showing the difference in miles between actual citations (mean) and the control group (control) in terms of their frequency. The dark brown represents overlap between these two ……… 27

Figure 5: The distribution of patent count for each cited patent in the dataset …………………………. 33

Figure 6: A histogram showing the breakdown of cited patents into periods of 1977-1984, 84-91, 91-98, 98-2005, 05-12…… 36

Figure 7: A histogram showing the lag time in years between originating patents and their citation, with a mean of 6 years…….. 36

Figure 8: This line graph shows the difference in percentage loss relative to the baseline for original patents with priority year before and after 1994 ………………………………………………………………….. 38

Figure 9: The percentage points decrease in probability with a lag time of less than 6 years, against a lag time of greater than 6 years ………………………………………………………………………………………… 41

Figure 10: The percentage point decrease in the probability of citation in the different industries where distance is significant, against the zero condition of 0-25 miles………………………………………… 42

List of tables

Table 1: *Descriptive statistics for geocoded data* ………………………………………………………………………….. 25

Table 2: *Binomial t-test results of matching percentage of citations against the control group* …. 30

Table 3: *Baseline Probit regression margin results* ………………………………………………………………………… 32

Table 4: *The effects of distance on citation probability across different periods* ………………………… 37

Table 5: *The effects of citation lag on patent citation* ………………………………………………………………… 40

Table 6: *The effects of geography across industries* …………………………………………………………….... 43

List of acronyms and abbreviations

MSA	Metropolitan Statistical Area
EPO	European Patent Office
PCT	Patent Co-operation Treaty
IPC	International Patent Classification
OECD	Organisation for Economic Co-operation and Development

1) Introduction

The economics literature has sought to explain why we see the agglomeration of firms since the workings of Alfred Marshall on the three benefits of agglomeration in 1890 (McCann, 2013). This has encompassed the development of a wide variety of models and theories following on from Marshall's initial ideas, including those from different disciplines, that have sought to understand the factors that result in agglomeration/cluster formation. The fact that these theories and models continue to be developed and commented on highlights the importance ascribed to this area of research in terms of being able to understand how cities and regions can support agglomerations and hence economic growth. The issue with this, however, is that the considerable number of theories and commentaries has led to confusion as to what an agglomeration/cluster is and how it can be measured (Parr, 2002). In this sense, one consistent strand that runs through most of these theories is the idea of knowledge spillovers, which can be explored to further our understanding of the concept of agglomeration and hence ultimately support economic growth.

This idea of knowledge spillovers was originally conceived by Alfred Marshall as 'The mysteries of the trade become no mysteries; but are as if it were in the air' (Marshall, 1890, p. 251), which is similarly reflected in subsequent theories. However, its importance and dynamic understanding can be more readily seen in the recent developments of the agglomeration literature, such as in Endogenous Growth Theory and Porter's Cluster concept (Fritsch & Franke, 2004). Here, knowledge and knowledge spillovers have become recognised as important determinants of innovation and drivers of local and regional economic growth. Therefore research has sought to understand the factors which can affect this. This has included the use of wage rates, worker mobility, industry concentration, the flow of goods and other methodologies to explore how geography can influence knowledge spillovers. However, recent research in this area has utilised patent and patent citations to proxy for knowledge spillovers due to the highly granular location information they contain (Belenzon & Schankerman, 2013; Buzard, et al., 2020), along with their correlation with communication efforts between firms and their relationship with innovative outcomes (Jaffe, et al., 2000; Duguet & MacGArvie, 2005).

Therefore, in this dissertation, how distance and administrative boundaries affect knowledge spillovers is explored using patent citation data from the OECD REGPAT database. This extends the existing literature by examining the existence of knowledge spillovers within the UK, whereas most previous research has focused on the example of the United States (Howells, 2002). This also includes questions such as how this relationship has changed over time, how knowledge diffuses over time and how knowledge spillovers operate in different industries. To this end, initial exploration is performed using a binomial t-test, in line with previous research, while a Probit regression methodology allows for clearer attribution to the effect of geography as both distance and administrative boundary effects are examined simultaneously. The main aim therefore is to complement and extend the existing literature by exploring knowledge spillovers within the UK, while seeking to use the results and subsequent insights gained to help inform policy decisions that seek to foster the diffusion of ideas and knowledge between firms to support economic growth.

The results show that local (NUTS3), regional (NUTS1) and country boundaries are positively associated with knowledge spillovers, and that most knowledge spillovers are exhausted within a 100-500-mile range. Furthermore, the effects of distance and administrative boundaries have become stronger over the period of 1977-2019, and that distance no longer constrains knowledge diffusion after a period of 12 years. Finally, knowledge spillovers operate differently across industries, with some industries exhibiting strong distance-based effects, while others have weaker or even non-significant distance-based effects. These findings therefore have important implications for industrial and innovation policy that aims to generate economic growth through the transfer of knowledge between firms.

The history of agglomeration economies and knowledge spillovers are discussed in section 2 below. The data is initially explored and explained in section 3, while the methodology is presented in section 4. Section 5 presents the results of the effects of distance and administrative boundaries through the binomial t-test and Probit regression, with the discussion and implications of these results found in section 6. Finally, the conclusion is provided in section 7.

2) Literature review

As new communication technologies have been developed and global integration has increased, there have been claims that the role of location, and hence distance, is decreasing in importance (Martin & Sunley, 2003). This would suggest that agglomerations of firms should become less prevalent because they result in higher land prices and congestion due to competition and increased resource use, therefore increasing overall production costs (McCann, 2013). However, as Glaeser (2010, p.1) notes, 'a paradox of our time is that … industrial agglomerations remain remarkably vital'. For example, in Denmark and Sweden manufacturing firms have become more geographically concentrated (Maskell & Malmberg, 1995), and in the US, innovation is concentrated in certain metropolitan areas and regions such as Silicon Valley and New York (Buzard, et al., 2020). Therefore, it has been suggested that economists should not only ask 'what to produce, how to produce, and for whom to produce', but also 'where to produce' (Feldman & Kogler, 2010, p. 404). Consequently, it is important to understand the mechanisms through which location can affect these economic outcomes.

The first theory relating to the concentration of firms comes from Adam Smith in 1776 in relation to the internal economies of scale (McCann, 2013). This argues that firms should invest in a larger factory rather than dispersed production facilities as this would allow for specialised workers and fixed costs to be spread across more output, leading to increased worker productivity, lower costs and hence greater profitability (Bathelt, et al., 2004). However, this theory only explains the existence of large individual factories rather than the concentration of multiple firms.

Building on from this therefore was the idea of external economies of scale. This developed from three observations that: a few concentrations of firms produced a large number of the worlds manufactured goods, firms in similar or related industries were often seen located in the same place, and that both of these observations were consistent through time (Malmberg, 1996). Marshall, in his *Principles of Economics* (1890), was the first to develop these observations into a consistent theory based on three reasons of: a local skilled labour pool, knowledge spillovers and non-traded local inputs (Martin & Sunley, 2003). The first of these suggests that a concentrated group of firms attracts a large pool of highly skilled workers, providing a firm with more productive matches with workers (Helsey & Strange, 1990), and allowing the firms to react quicker to changes in the market through hiring and firing decisions (Puga, 2010). Secondly, knowledge spillovers were suggested to result from the increased access that workers and firms have to each other due to close geographical proximity, allowing for more frequent and informal face-to-face interactions and hence the transfer of knowledge (O'Sullivan, 2012). Finally, non-traded local spillovers benefits firms that require intermediate inputs that are difficult to transport as a greater number of firms increases demand for these inputs which allows the intermediate firms to specialise and hence increase productivity (Puga, 2010). Thus, local industrial growth would result in cost savings to the firms within the 'industrial cluster', in excess of the increases in costs associated with concentration, leading to the agglomeration of firms (McCann, 2013).

These ideas continued to be expanded on and experimented with throughout the 1940s and 50s, including the distinction between localisation and urbanisation economies coming from the work of Ohlin (1933), Hoover (1948), and Isard (1956) (McCann & Van Oort, 2019). Here, the idea of localisation economies is where firms benefit primarily from the expansion of their own industry within the local area, regardless of the size of other local industries (Feser, 1998). For example, a large industry would shape local resources to its needs, allowing it to grow faster than average in comparison to other local industries (McCann, 2013). Alternatively, urbanisation economies suggested that all local firms would benefit from local growth, regardless of the industry (Feldman & Kogler, 2010). For example, individuals may be able to specialise further as a result of matching opportunities across different local industries due to greater demand and reduced risk, thereby increasing productivity (Rosenthal & Strange, 2004). Jacobs (1961) also subsequently pointed out the potential of urban diversity, in contrast to localisation effects and extending the ideas of urbanisation. This suggests that a diverse range of firms within an urban agglomeration can boost creativity, due to the potential for the cross-fertilisation of ideas across industries (Feldman & Kogler, 2010).

However, these theories were criticised for their emphasis on transaction costs, with firms being able to costlessly change their relationships with others, and the fact that the benefits of concentration were simply seen as operating because they were geographically localised (Maskell, 2001). Therefore, these theories ignored the effects of socialisation within the process of production and knowledge transfer, including the history, size and institutions of a local community (Harrison, 1992). However, with the revival of agglomeration economies in the 80s and 90s, theories began to acknowledge the role of dynamic influences such as social, cultural, and political factors that play a role in cluster formation, development, and subsequent decline (Malmberg, 1996). Consequently, simple infrastructural cost advantages were no longer seen as the only reason why firms would concentrate (McCann, 2013).

This re-emergence and change in focus occurred in the 1980s and 90s due to a shift towards institutional and evolutionary issues within economics and the recognition of the significant fall in transport costs during the 20th Century (Malmberg, 1996). This meant that infrastructural cost factors were no longer as important as they were before (Malmberg, 1996). Furthermore, it was also acknowledged that knowledge and knowledge intensive sectors were becoming increasingly important for driving economic growth, such that these had to be integrated into the development of new models (McCann & Van Oort, 2019). This therefore led to several new theories and concepts, including the development of New Economic Growth Theory and Endogenous Growth Theory which provided a mathematical basis for clustering (McCann & Van Oort, 2019). These were developed following the Dixit-Stiglitz monopolistic competition model, allowing for increasing returns to be included within economic models (Krugman, 1998), with Endogenous Growth Theory explicitly accounting for knowledge spillovers for increasing returns (Carlino & Kerr, 2015). Furthermore, evolutionary and institutional approaches developed which stressed the importance of institutions and policy-makers in cluster formation and growth (McCann & Van Oort, 2019). Here, the emphasis was on the way in which institutions allow for the transfer of knowledge between firms, which was seen as crucial for innovation in knowledge intensive sectors (Speldekamp, et al., 2020). Finally, several different conceptions of agglomeration evolved from different strands of literature, including from neo-classical economics, the business and management literature, geography and spatial planning, and the sociological literature (Gordon & McCann, 2000).

Therefore, new and existing literature exploring the topic of agglomeration economies has produced many different concepts with different emphases. This has consequently led several authors to comment on the confusion this has generated in terms of defining what a cluster or agglomeration is and how it may be measured (Parr, 2002; Martin & Sunley, 2003). However, all these theories are consistent in arguing that agglomeration, and the factors that affect it, are crucial for regional and urban performance, therefore research must continue to explore and understand this concept. In this sense, one of the more consistent factors found in these theories is that of knowledge generation, accumulation and spillovers, which are suggested to be one of the key forces driving clustering benefits (McCann & Van Oort, 2019). Specifically, theories that have developed since the 1980s highlight the fact that there has been a shift in emphasis towards recognising the importance of knowledge spillovers and localised interaction leading to both learning and innovation (Malmberg & Maskell, 2002). This is in line with the acknowledgement by both academics and policy makers that knowledge intensive industries and sectors are crucial for modern economic growth (McCann & Van Oort, 2019). Therefore, while it may be difficult to separate out and identify the different factors that may affect agglomeration economies (O'Clery, et al., 2019), the emphasis on knowledge as a driver of agglomeration and economic growth through innovation suggests that it is important to understand the mechanisms and extent of these knowledge spillovers.

The initial idea of knowledge spillovers appears in Marshall's three benefits of agglomeration, wherein which knowledge invested in by one firm or individual can spillover to other firms if they don't have to invest the same resources, as long as they are located in close proximity (Jaffe, et al., 2000). This process was originally thought to occur costlessly, but it was not until the re-emergence in the 1980s and 90s that it was recognised that individuals and firms had to invest resources and time into the community to benefit from this exchange (Malmberg & Maskell, 1997). This was compounded by the suggestion that external sources of knowledge are important for firms to remain at the forefront of innovation, therefore emphasising the role of external knowledge exchange in generating new ideas and technologies (Mackinnon & Cumbers, 2011). Thus, agglomeration of firms were seen to occur due to the distinction between tacit and codified knowledge (Polanyi, 1958). This is because, while both forms of knowledge are required for innovation (Charlot & Duranton, 2006), tacit knowledge is less formalised and requires face-to-face interaction to be transferred (Howells, 2002). Hence, the benefit of colocalization is the potential for repeated face-to-face interaction and serendipitous transfer of tacit knowledge, leading to innovation (Charlot & Duranton, 2006).

In this regard, there is a significant corpus of literature that has attempted to explore the different factors that may affect this, utilising different methodologies and conceptions of how knowledge spillovers may be measured. This is primarily because of the difficulty of measuring knowledge spillovers, as emphasised by the following quote from Krugman (1991, p.53) that "knowledge flows … leave no paper trail by which they may be measured and tracked". Therefore, any method that attempts to do so must measure them indirectly (Thompson, 2006). One of the ways in which this has been explored previously has been through the impact of concentration on learning and economic development, as measured through wages (Glaeser, 1999; Glaeser & Mare, 2001), or through tracking innovative individuals through their movements between firms (Breschi & Lissoni, 2000; Miguelez & Moreno, 2015). The issue with these methodologies, however, is being able to collect the data to identify pure knowledge spillovers. This is because learning and moving between firms can be mediated through the market mechanism, making it a pecuniary externality rather than a pure agglomeration externality (Henderson, 2007). Alternatively, another method argues that knowledge spillovers can be measured through the trade of goods between firms and regions (Kogler, 2015). This is because licences or goods purchases can be used to identify new technology or techniques that can subsequently be built on by the firm, to produce new innovative outputs (Kogler, 2015). However, these could be classed as 'rent spillovers' as they are captured within the market mechanism, thus not representing pure knowledge spillovers, and data available to study this phenomenon is mostly at the highly aggregated country level, restricting the granularity of these studies (Feldman & Kogler, 2010). Therefore, the most common method of studying knowledge spillovers is that of inputs and outputs of innovation, including R&D spending, literature-based innovations and patenting (Carlino & Kerr, 2015). This is primarily because of the granularity of data available, along with the fact that knowledge spillovers can often be most easily tracked through these methods (Feldman & Kogler, 2010).

In this sense, the innovation-based measure that is most common to explore knowledge spillovers is that of patent and patent citations. Their use for this purpose has followed the digital publication of patents and their citations in the late 1990s (Carlino & Kerr, 2015). They are part of the broader legal intellectual property framework whose purpose is to grant an inventor a legal monopoly over their invention for a given amount of time (Hettinger, 1989). A firm or individual can be granted a patent when the invention is new, not obvious, and capable of industrial application, suggesting that they must have some economic value before they can be patented. Furthermore, the application process is often costly and time consuming, therefore their value must exceed the cost of the application (Moreno, et al., 2005). Hence, they can be used as a proxy for innovation because of their economic value and the fact that they must represent a new idea or technology in order to be granted (Carlino & Kerr, 2015).

In terms of knowledge spillovers, patents can be used as a proxy for the paper trail of knowledge as they contain information on, including the location of, the firm and the inventor, along with previous patents that they may be related to (Thompson & Fox-Kean, 2005). This is because the citations must ensure there is no infringement of previous patents and so that the new innovation can be clearly demarcated (Jaffe & Trajtenberg, 1996). Evidence supporting this claim comes from Jaffe et al. (2000) and Duguet and MacGarvie (2005) who find evidence of communication between citing and cited patent inventor teams along through informal communication or technology transfer. Furthermore, based on the results from the US, evidence suggests that potential knowledge flows embedded in citations are geographically localised (Belenzon & Schankerman, 2013; Singh & Marx, 2013; Buzard, et al., 2020). Thus, patents have often been used to explore factors that affect knowledge spillovers, such as distance, administrative boundaries, institutional factors and time.

However, issues of using this measure must also be recognised. This includes the fact that not all industries patent making it difficult to generalise the results (O'Clery, et al., 2019), not all inventions that can be patented are, due to the requirement of full disclosure (Agrawal, et al., 2014), they represent codified knowledge which can be more easily transferred without interaction (Howells, 2002), and that propensity to patent may change over time (Sonn & Stoprer, 2008). Despite this, they are accepted as one of the best mechanisms for measuring knowledge spillovers and thus widely used in academic research (Buzard, et al., 2020).

In most studies utilising this methodology or exploring the concept of knowledge spillovers within agglomeration, the literature has tended to focus on the example of the United States (Howells, 2002; Belenzon & Schankerman, 2013; Singh & Marx, 2013; Buzard, et al., 2020). To the knowledge of this author, those that do examine the case of the United Kingdom have so far explored agglomeration more widely, such as the geographic distance over which firms in different industries may cluster (Duranton & Overman, 2005), or at a much smaller scale of a single science park using patent citations (Helmers, 2019). Therefore, this paper seeks to extend this corpus of literature by examining the case of the UK more thoroughly in terms of the effects that geography may have on knowledge spillovers. The purpose of this is to be able to compare the results to those of the United States and suggest the ways in which policy may use UK evidence to support or change existing innovation policy, without relying on results generated in the US.

In doing so, several different questions are explored for the purpose of aiding policy, relating to the role of administrative boundaries, distance, changes over time, knowledge diffusion and differences across industries. Firstly, in terms of the role of administrative geographical units, such as the metropolitan statistical area, regions and country level, early research using patents used these to proxy for distance, finding statistically significant results of bounded knowledge spillovers within each of these areas (Jaffe, et al., 1993). Subsequent research since has also supported this finding (Belenzon & Schankerman, 2013; Singh & Marx, 2013). However, beyond being a proxy of distance this effect is suggested to also be because of institutions, such as informal codes of interaction, technical language usage and worker mobility, that can support the process of knowledge spillovers (Malmberg, 1996). This is because these institutions are suggested to be constrained within local, regional or country level administrative boundaries (Agrawal, et al., 2014; Speldekamp, et al., 2020). As such, administrative units are suggested to constrain knowledge spillovers not just because of the effects of distance, but also role that institutions may play within them. It is therefore reasonable to assume that such findings are also to be expected to be found in the UK, despite differences in scale at the local, regional and country level. Hence, the first hypothesis is that:

> *Hypothesis 1:* Administrative boundaries will have a positive and significant effect on the probability of citation at the local, regional and country level in the UK.

The second question builds on this by seeking to explore the geographic extent over which knowledge spillovers operate in the UK. This is a key question that is still being examined in the case of the United States with recent research finding evidence of knowledge spillovers up to a distance of 50 to 150 miles (Belenzon & Schankerman, 2013; Singh & Marx, 2013). Therefore, based on the distinction between tacit and codified knowledge, with the former relating to knowledge spillovers primarily through face-to-face interaction, this would suggest that they operate over an extended commuting boundary (Belenzon & Schankerman, 2013). However, research exploring knowledge spillovers at a more local level within MSAs has found evidence that knowledge spillovers are primarily exhausted within a distance of 20 miles (Buzard, et al., 2020), or even within a few hundred metres (Helmers, 2019). In this sense, it must be recognised that the finding of these forces of agglomeration can be influenced by the scale of analysis (Cottineau, et al., 2019). Therefore, given the use of the NUTS3 classification or equivalent as the location assigned to individuals, it would be expected to find results that are congruent with those suggesting knowledge spillovers operate within a scale of 50-150 miles. Thus, we get hypothesis 2 that:

> *Hypothesis 2:* Most of the knowledge spillovers will be exhausted within a range of 0-150 miles.

However, several other questions must also be asked for the purpose of policy discussions. This includes how the effect of distance has changed over time on knowledge spillovers (Charlot & Duranton, 2006). This is particularly important given the rise of new communication technologies that many have suggested could lead to the 'death of distance' given their ability to facilitate regular long distance communication (Sonn & Stoprer, 2008). If, therefore, technology can support this communication and hence tacit knowledge sharing over longer distances, replacing face-to-face interaction, then we would expect to see the effects of both distance and administrative boundaries decreasing over time. Such results would be especially important given the rise of work from home due to the Coronavirus epidemic and predicted continuance of this trend (Hern, 2020; Eisenberg, 2020; Regan & Harby, 2020). Therefore, this leads to hypothesis 3 that:

> *Hypothesis 3:* The strength and importance of administrative boundaries and geographic distance in knowledge spillovers will decrease over time.

This can also be related to the suggestion that knowledge diffuses over time, such that the longer a patent is available, the less likely distance is to affect its probability of citation (Verspagen & Schoenmakers, 2004). This is in line with many other theories that consider the diffusion of technology, and that its adoption increases over time as it geographical reach extends further and is introduced into more and more marginal communities (Geroski, 2000). Hence, this leads to hypothesis 4 that:

> *Hypothesis 4:* The longer the lag between a patent and its citation, the weaker the role of distance or administrative boundaries in its dissemination.

Finally, it is also important to explore the different effects that distance and administrative boundaries may have across different industries (McCann & Van Oort, 2019). This is because there is a significant corpus of literature that looks at agglomerative forces, and knowledge spillovers, across different industries and finds different results. In terms of the UK, Charlot and Duranton (2006) examined industry location and found that while for some industries localisation was important at scales of 0-50km, others were more significant at regional scales of 80-140km. Extending this to the use of patents for examining knowledge spillovers, Belenzon and Schankerman (2013) in the US found that the effects of distance are less pronounced in Biotechnology, IT, and Telecommunications compared to Chemical and Pharmaceutical industries. In contrast, Adams and Jaffe (1996) found that spatial proximity was not significant in the Pharmaceutical R&D industry. This could be related to the different propensity to patent across industries (O'Clery, et al., 2019), however, the fact that different results are found for similar industries suggest there is still some ambiguity. Hence, this must be explored in the case of the UK so that policies could adjust depending on the industry they are dealing with. Therefore, the final hypothesis is that:

> *Hypothesis 5*: Knowledge spillovers will be affected differently across industries by administrative and distance boundaries.

3) Patent data exploration

Data for this paper comes from the OECD REGPAT database, containing information on patents applied for under the European Patent Office (EPO) and the Patent Co-operation Treaty (PCT) (OECD, January 2020)[1]. This includes the patents application number, original application year and International Patent Classification (IPC), along with the individual's/firm's address, country code and share of the patent application for inventors/applicants. From this, patents whose entire inventor team was located in the UK were identified because with international teams it becomes difficult to identify from which country the innovation originated. Therefore, given the focus on UK knowledge spillovers only UK inventor teams were used (Belenzon & Schankerman, 2013). Furthermore, for identifying the location of the patent the individual inventor location was used, rather than the firm's, because firms often register patents at their headquarters which creates ambiguity as to where the patent was actually developed (Moreno, et al., 2005). While the use of this type of location information may limit the exploration of clusters of firms it provides a more reasonable estimation of the scale over which knowledge spillovers occur which can then more clearly inform the scale over which clusters may operate (Buzard, et al., 2020). Hence, 240,045 patents were identified that matched these criteria from the period of 1977 to 2019. The distribution of these patents according to the share of the patent assigned to the inventor and hence their local NUTS3 boundary can be seen in figure 1 below. This shows that while inventors were distributed across all regions in the UK, the creation of patents is clustered in the South East.

[1] The latter application process is an international one that, while not granting any legal protection, allows applicants to subsequently apply to regional patent offices including the EPO. For this paper, only patents that subsequently apply to the EPO, not already included in the EPO dataset, are extracted using the PCT-EPO correspondence table (OECD, January 2020).

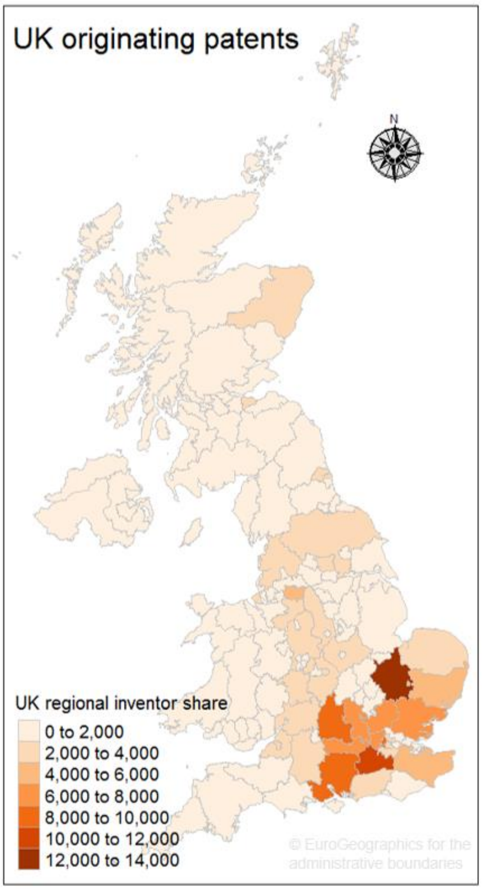


Figure 1 – The local (NUTS3) sum of patent inventor share for individuals from 1977 to 2019 whose entire inventor team is located in the UK. This refers to the share of the patent assigned to the individuals when the patent is applied for.

Information on patent citations was then subsequently obtained from the OECD patent citations database (OECD, January 2020). This contained information on citations made under both the EPO and the PCT, including the original and citing application number[2]. Consequently, of the original 240,045 patents identified, 45,852 had been cited, with the regional distribution of these seen in figure 2 below.

[2] The PCT dataset was used separately from the EPO one due to citation searches for patents that originally applied under the PCT occur within the PCT database, not when they are subsequently registered under the EPO (Webb, et al., 2005). Therefore, information about citation pairs were obtained through both the PCT and EPO datasets separately and later merged.

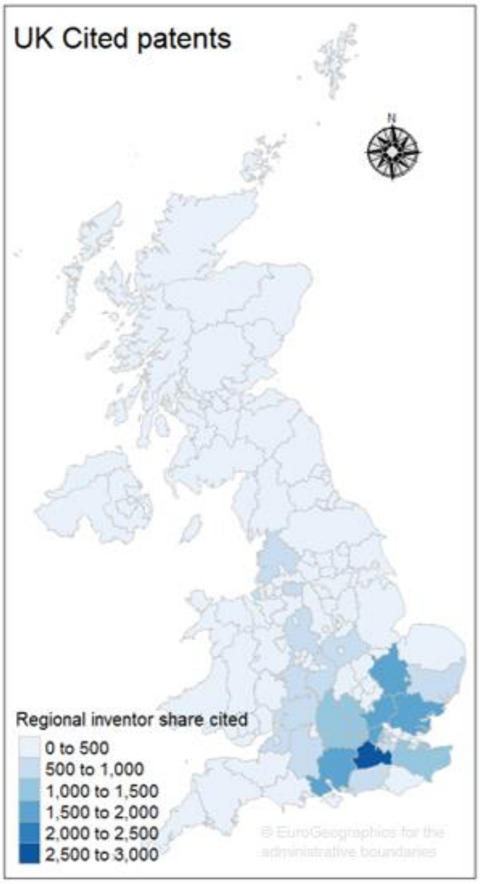
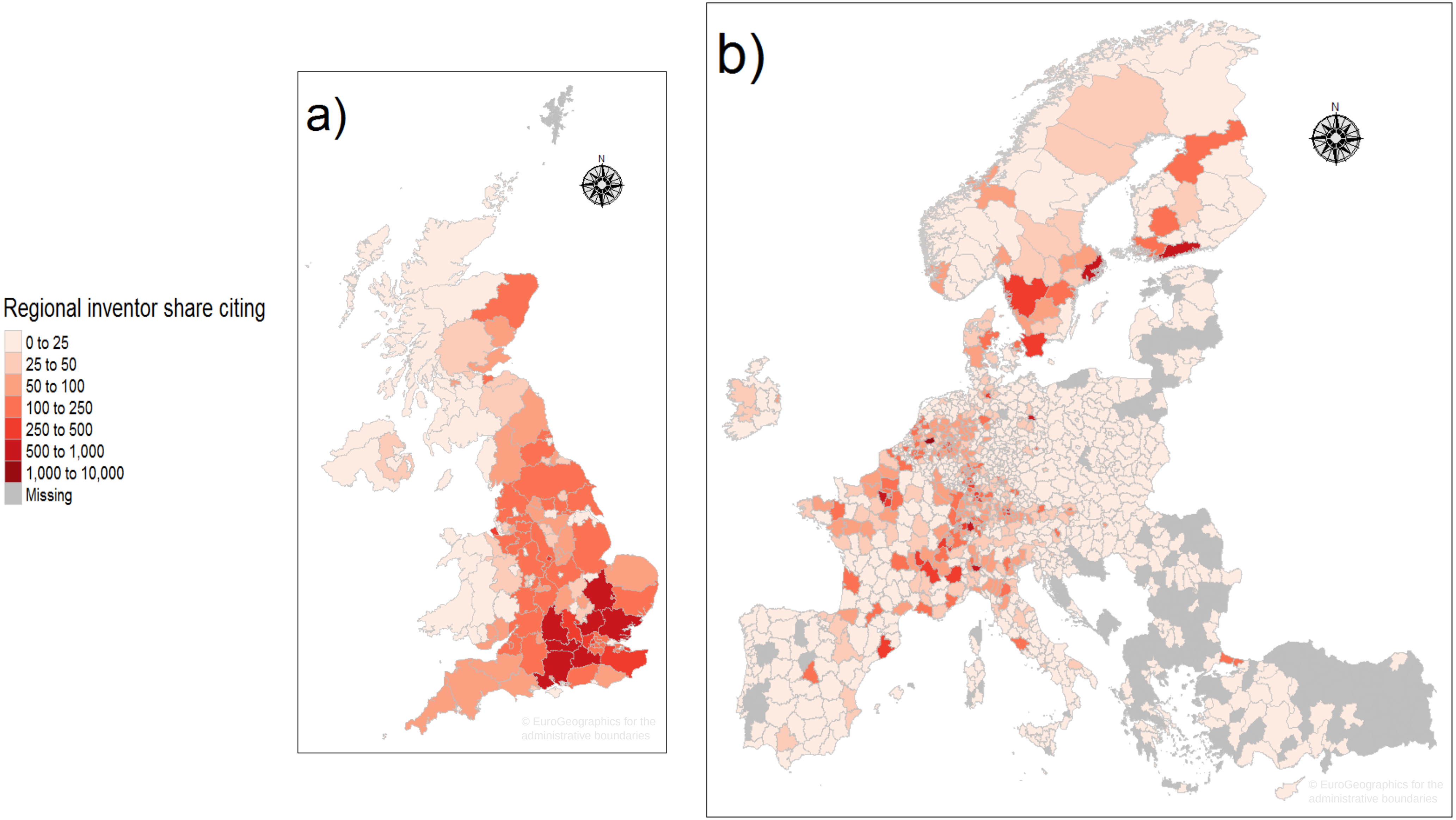
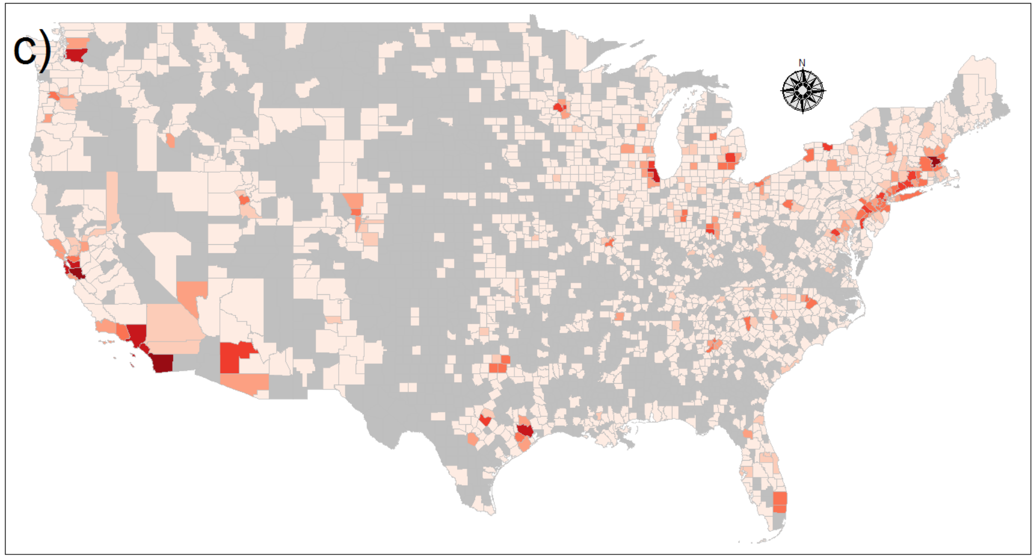


Figure 2 - The regional distribution of the sum of patent inventor share by individuals from 1977 to 2019 whose entire inventor team is based in the UK and whose patent has been cited by applications under the PCT or the EPO.

Subsequently, self-citations by firms or individuals were removed from this dataset because they do not represent knowledge spillovers as there is no external transfer of information (Jaffe, et al., 1993). Furthermore, individual locations were matched with local centroid values at the NUTS3 (local) level or equivalent, resulting in 43,793 original patents remaining, with 145,633 cited-citing pairs. These locations were based on the NUTS3/LT3/county level scale that the OECD REGPAT database assigned to individuals according to their address (OECD, January 2020)[3]. Most of these citations came from areas such as the UK, Europe, America, China, Japan and South Korea. Figure 3 below highlights this by showing the regions from which the most inventors of actual citations could be attributed to, with many of these reflect key areas of innovation across the globe.

[3] To see the regions that were covered and how these were matched see Appendix A.


Regional inventor share citing
0 to 25
25 to 50
50 to 100
100 to 250
250 to 500
500 to 1,000
1,000 to 10,000
Missing
a)
N
© EuroGeographics for the administrative boundaries
b)
N
© EuroGeographics for the administrative boundaries



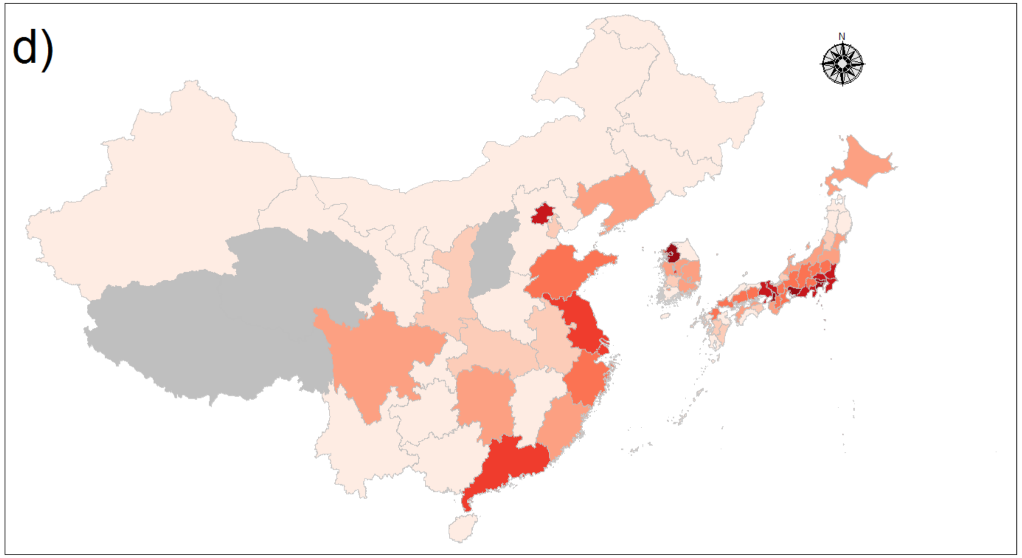


Figure 3 - The regional distribution of the sum of patent inventor share by individuals from 1977 to 2019 from patents that cite UK patents. The regions that produce the most citations are shown here with a) UK, b) Europe, c) United States, d) China, South Korea and Japan.

Table 1 - Descriptive statistics for geocoded data

		Combined		EPO		PCT	
		%	count	%	count	%	count
Count		100%	145,633	46.99%	68,426	53.01%	77,207
Local		2.84%	4,140	5.87%	4,018	0.16%	122
Regional		4.54%	6,607	8.51%	5,822	1.02%	785
Country		9.99%	14,556	15.56%	10,647	5.06%	3,909
0-25 miles		2.44%	3,559	5.00%	3,419	0.18%	140
25-50 miles		1.16%	1,690	1.95%	1,333	0.46%	357
50-100 miles		1.78%	2,593	2.61%	1,783	1.05%	810
100-250 miles		5.09%	7,416	6.84%	4,682	3.54%	2,734
250-500 miles		13.70%	19,950	17.36%	11,876	10.46%	8,074
500-1000 miles		17.83%	25,967	20.20%	13,824	15.73%	12,143
1000-2500 miles		4.49%	6,539	3.57%	2,446	5.30%	4,093
2500-5000 miles		25.59%	37,262	21.50%	14,713	29.21%	22,549
5000+ miles		27.92%	40,657	20.97%	14,350	34.07%	26,307
Mean distance (miles)	Mean		**2922.88**		**2341.03**		**3438.55**
	Median		**3293.20**		**746.47**		**3859.41**
Minimum distance (miles)	Mean		2809.72		2234.71		3319.32
	Median		3215.95		657.13		3696.92

This table presents the percentage and count of patents citations within each administrative and distance boundary within the EPO, PCT and combined relative to the patent they cited.

Once individual locations were matched to localities at the NUTS3 scale, distance could then be calculated using the Haversine formula based on the latitude and longitude values[4]. For this, the mean distance between patent teams were used, such that if multiple inventors were involved in a patent, all distances were calculated between them and then averaged. This is because it is not clear as through which connection knowledge may have travelled and thus the average distance between teams must be used (Belenzon & Schankerman, 2013). Administrative boundaries could then also be defined as the local (NUTS3), regional (NUTS1) and country scale. These results in terms of the percentage of patents within administrative and distance boundaries are therefore shown in table 1 above.

[4] This comes from the Haversine library in python, with distance calculated 'as-the-crow-flies'.

From this it can be seen that many of the citations originate beyond distances which we would expect to see knowledge spillovers occurring as the mean and median for all three datasets are in the range of 700-4000 miles. However, much of this simply reflects the existing distribution of industry and are therefore from where we would expect patents to originate. To control for this and to show that knowledge spillovers are localised compared to the distribution of existing industry (Duranton & Overman, 2005), a control sample of patents is created (Murata, et al., 2014). This control sample is based on each control patent matching on the 4-digit IPC code and priority year of the citing patent, but that did not cite the original patent (Thompson & Fox-Kean, 2005). From there, a random control patent was extracted for each original/citation pair, to be able to compare the actual distribution of knowledge spillovers against the potential distribution of industrial activity (Carlino & Kerr, 2015). This therefore follows the matching methodology developed by Jaffe et al., (1993) which has subsequently been used by many papers in exploring knowledge spillovers to control for the existing distribution of technology creation (Kogler, 2015).

The distribution of the actual citation and control dataset in terms of distance can be subsequently seen in figure 4 below. This shows that in the range of 0-150 miles that we would expect knowledge to see knowledge spillovers operating, actual citations far outweigh control citations, suggesting evidence of localisation. Furthermore, spikes of overall citations/control patents can be seen within ranges of 0-1000 miles, 3000-4000 miles, and 5000-6000 miles corresponding to citations/patents that originate within the UK and Europe, the United States of America, and Eastern China, Japan and South Korea respectively, with the final spike coming from Australia and New Zealand.

A complete workflow for this can be found in Appendix B, along with the link to the GitHub repository.

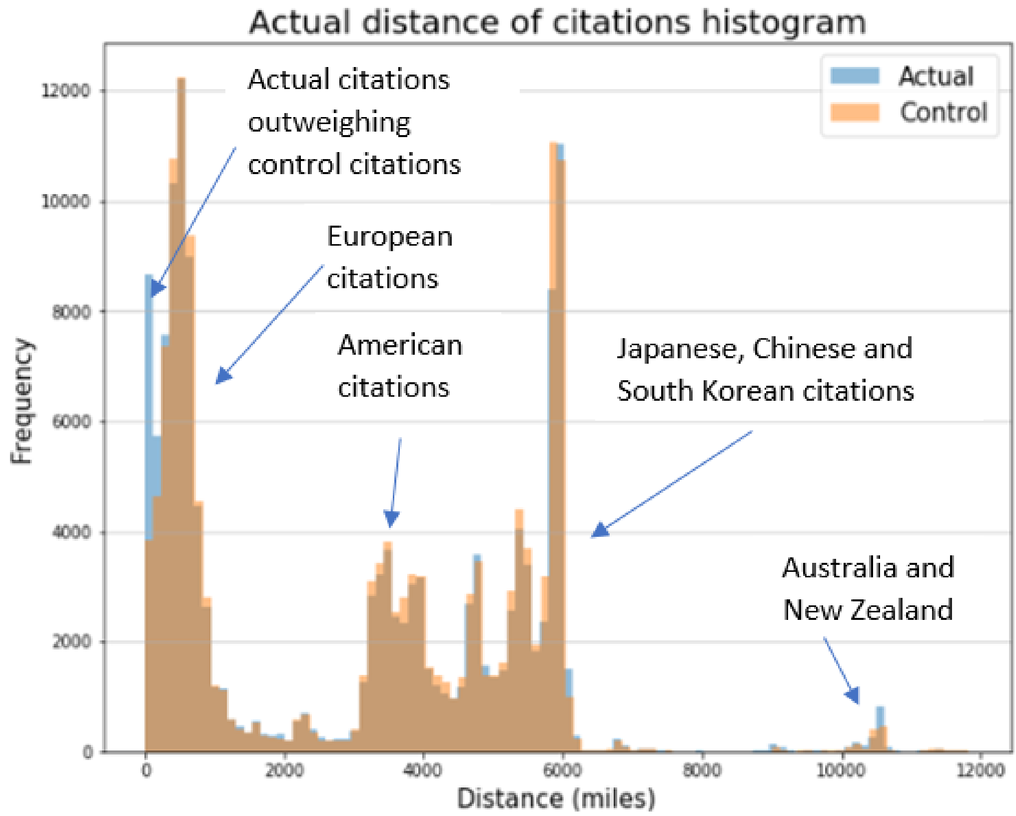


Figure 4 - A histogram showing the difference in miles between actual citations (mean) and the control group (control) in terms of their frequency. The dark brown represents overlap between these two.

3.1) Statement of ethics

The use of patents includes identifying and geographically assigning individuals to regions that are associated with specific patents. This however can be used because of the requirement of disclosure when patents are applied for, and the information is readily available in a public database. Therefore, no ethical concerns were raised in the development of this project.

4) Methodology

This data is then used to test for the hypotheses detailed in the literature review, to examine how knowledge spillovers, as measured through patents, operate in the UK. Firstly, to test for localisation at the local (NUTS3), regional (NUTS1) and country level for hypothesis 1, a binomial t-test is used (Jaffe, et al., 1993). This is based on whether one of the regions assigned to the original and citing/control patents are the same. Therefore, this uses the following formula:

$$t = \frac{P_c - P_o}{\sqrt{(P_c(1 - P_c) + P_o(1 - P_o))/n}} \quad (1)$$

The hypothesis to test is that H_0: $P_c = P_o$ against H_a: $P_c > P_o$ where P_c is the proportion of actual citations that come from the same geographical unit that the original patent comes from, while P_o is the proportion of control citations that come from within the geographical boundary (Jaffe & Trajtenberg, 1996). The t-value is then used to calculate the significance of this result, which would suggest whether there is evidence of localisation at these scales.

However, there are several issues with this being used to examine evidence of localisation. Firstly, adjacent administrative units are treated the same as units that are at the opposite end of the country and therefore may underestimate the effect of knowledge spillovers (Murata, et al., 2014). Secondly, it can only test for one boundary condition at a time (Singh & Marx, 2013). Thirdly, administrative boundaries do not always reflect economic boundaries and therefore distance must also be considered (Duranton & Overman, 2005). Subsequently, to explore fully all hypotheses in controlling for both boundary conditions and distance simultaneously a regression model is used.

For this, a Probit regression specification was chosen given the dichotomous dependent variable of whether the control/citing patent did cite the original patent[5]. An OLS specification in this case suffers from heteroscedasticity and does not restrict the dependent variable to be within the range 0 or 1 which would lead to inconsistencies and lack of interpretability. Therefore, a non-linear approach must be used to restrict the value of the dependent variable between 0 and 1 and control for these errors (Allison, 2017). Furthermore, a Probit regression was chosen rather than a logistic regression because of its prevalence within econometric literature as opposed to the logistic model (although the results are robust to the use of a logistic model as seen in Appendix C). This Probit regression therefore follows the specification of:

$$P(Y_{ij} = 1 | D_{ij}, D_{wb}, T_{ij}) = \Phi(\beta_0 + \beta_1 D_{ij} + \beta_2 D_{wb} + \beta_3 T_{ij} + \epsilon_{ij}) \quad (2)$$

[5] 1 = there is a citation, 0 = there is no citation

Where Y_{ij} is a dummy variable equal to 1 if patent i is cited by patent j and 0 otherwise, D_{ij} represents dummy variables for sets of distances between i and j, D_{wb} are a set of border dummy's for if patent i is within the same administrative border as patent j, and T_{ij} is a dummy variable indicating if there is a match at the 6-digit IPC level between i and j. Furthermore, ϵ is given as the error term and Φ represents the Probit link function (Talibova, et al., 2018). This therefore controls for distance and boundary effects, along with a greater precision of technology matching given the critique of Thompson and Fox-Kean (2005) that knowledge is more likely to spillover within technology subclasses than between them, thus requiring a greater degree of specification of patent class.

The issue with this, however, is that there is no meaningful economic interpretation of the coefficients of a Probit regression (Jaffe, et al., 1993). This is because they represent the effect on the z-score for a one-unit change in the value of the independent variable and hence do not represent any economic quality (UCLA: Statistics Consulting Group, 2020). However, the margin values can be calculated for which the coefficient represents the percentage point change in the probability that the dependent variable would equal one if, *ceteris paribus,* the independent dummy variable changes from zero to one (Leeper, 2018). This is therefore performed for the regressions below.

5) Results

5.1) Binomial t-test results

The first test to understand the extent to which knowledge spillovers operate within the UK is that of the Binomial t-test. The purpose of this is to test for the evidence of localisation at the local (NUTS3), regional (NUTS1) and country level individually in relation to hypothesis 1. This does so by controlling for the existing distribution of industry through the control group such that if the actual citation probability of matching within an administrative unit exceeds that of the control group then this suggests that knowledge spillovers are localised. The results of these can be found in table 2 below.

Table 2 – Binomial t-test results of matching percentage of citations against the control group

	Administrative boundary	Count	Actual citation match (%)	Control citation match (%)	Difference in percentage values	T-statistic
EPO						
	local	136,852	**5.87%**	**0.63%**	**5.24%**	**(55.28)*****
	regional	136,852	8.51%	1.61%	6.90%	(58.93)***
	country	136,852	15.56%	6.57%	8.99%	(53.59)***
PCT						
	local	154,414	0.16%	0.14%	0.01%	(0.72)
	regional	154,414	1.02%	0.83%	0.19%	(3.92)***
	country	154,414	5.06%	4.81%	0.26%	(2.31)***
Combined						
	local	291,266	2.84%	0.37%	2.47%	(53.27)***
	regional	291,266	4.54%	1.20%	3.34%	(54.31)***
	country	291,266	9.99%	5.63%	4.36%	(43.99)***

*This table reports the results of a binomial t-test result comparing the percentage matching within the actual citation group against the control group at the local, regional and country level. This is done for the EPO, PCT and combined datasets. The significance level is indicated by * with * representing 10%, ** 5% and *** 1%.*

Here the results present evidence that knowledge spillovers are localised within each administrative boundary, supporting hypothesis 1. For example, in the EPO dataset the percentage of actual citations that have at least one inventor in the same region as an inventor from the original patent is 5.87%. In contrast, for the control group this percentage falls to 0.63% showing a difference in percentage values of 5.24 and that this difference is significant at the 1% significance level. This therefore presents evidence of localisation of knowledge at the NUTS3 administrative level in the EPO dataset as the actual citation percentage is significantly higher than the distribution of existing knowledge. This result is then replicated at the local, regional and country level in the PCT and combined datasets, except for the local level within the PCT dataset. The main significance of this is the evidence of the localisation of knowledge spillovers within the combined dataset which overall supports hypothesis 1. However, this does not control for actual distance between cited-citing pairs and is only able to test for localisation at one scale at a time.

5.2) Baseline regression estimation results

The baseline results for the regressions that explore the effects of distance and administrative boundaries simultaneously on the probability of citation can be found in table 3 below. Firstly, column 1 examines the effects of distance alone, in which the negative and significant coefficient suggests that as distance increases, the probability of citation decreases. However, this does not show the scale over which knowledge spillovers operate and thus is unable to consider hypothesis 2. Therefore, with the expectation that knowledge spillovers decay exponentially, the second column shows the effect of distance across different distance intervals. Within this specification 0-25 miles is given as the reference condition and so each distance interval is compared to this result. Here, the 25-50 mile boundary represents a reduction in the probability of citation by 25.7 percentage points against the 0-25 mile boundary, a 50% decrease against the baseline 0.500 probability of citation within the dataset. Moving then from 25-50 to the 50-100 miles interval leads to a further reduction of 7.0 percentage points (32.7-25.7) and the 100-250 miles results in a drop of another 3.2 percentage points (35.9-32.7). By the 250-500 mile boundary the total loss is 40.7 percentage points, after which distance has no more depreciable effect on citation. This is associated with an overall reduction of 80% against the 0.500 baseline probability of citation. Therefore, this suggests that knowledge spillovers can extend up to 500 miles, but that most of this effect is exhausted within 100 miles, supporting hypothesis 2.

Table 3 - Baseline Probit regression margin results

	Distance (1)	Distance boundaries (2)	Administrative boundaries (3)	Baseline (4)	Top quartile patents (5)	Bottom quartile patents (6)
Boundary match						
Local			**0.2668**	0.2167	0.1841	0.2265
			(0.000)***	(0.000)***	(0.000)***	(0.000)***
Region			**0.1218**	0.1061	0.1188	0.1241
			(0.000)***	(0.000)***	(0.000)***	(0.000)***
Country			**0.0541**	0.0363	0.0463	0.0287
			(0.000)***	(0.000)***	(0.000)***	(0.001)***
Distance (miles)						
ln(distance+1)	**-0.0211**					
	(0.000)***					
25-50		**-0.2567**		-0.0882	-0.032	-0.1143
		(0.000)***		(0.000)***	(0.472)	(0.000)***
50-100		**-0.3268**		-0.1081	-0.0132	-0.1112
		(0.000)***		(0.000)***	(0.752)	(0.000)***
100-250		**-0.3586**		-0.0935	-0.0153	-0.102
		(0.000)***		(0.000)***	(0.706)	(0.000)***
250-500		**-0.4072**		-0.1165	0.0026	-0.1435
		(0.000)***		(0.000)***	(0.949)	(0.000)***
500-1000		-0.4106		-0.118	-0.0058	-0.1325
		(0.000)***		(0.000)***	(0.886)	(0.000)***
1000-2500		-0.3825		-0.098	0.0671	-0.1692
		(0.000)***		(0.000)***	(0.100)*	(0.000)***
2500-5000		-0.4094		-0.1172	0.0351	-0.1724
		(0.000)***		(0.000)***	(0.384)	(0.000)***
>5000		-0.4151		-0.122	0.0169	-0.1739
		(0.000)***		(0.000)***	(0.676)	(0.000)***
Technology match						
IPC 6 match	0.1293	0.1263	0.1246	0.1238	**0.0752**	**0.1495**
	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***
R^2	0.0113	0.0148	0.0164	0.0166	0.0047	0.0313
log likelihood	-199610	-198890	-198580	-198530	-51139	-50139
LLR p-value	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***
Observations	291266	291266	291266	291382	74128	74670

*This table reports the coefficient estimates for the margin effects of a Probit regression model of citations to UK patents. These coefficients represent the expected percentage point increase in the probability of citation being 1, as a result of the dummy variable changing from 0 to 1. This focuses on the results when including administrative boundaries, distance boundaries against a baseline of 0-25 miles, and the quality of patents as measured by the number of citations. P-values are reported underneath the coefficients with statistical significance reported at the * 10%, ** 5% and *** 1% level.*

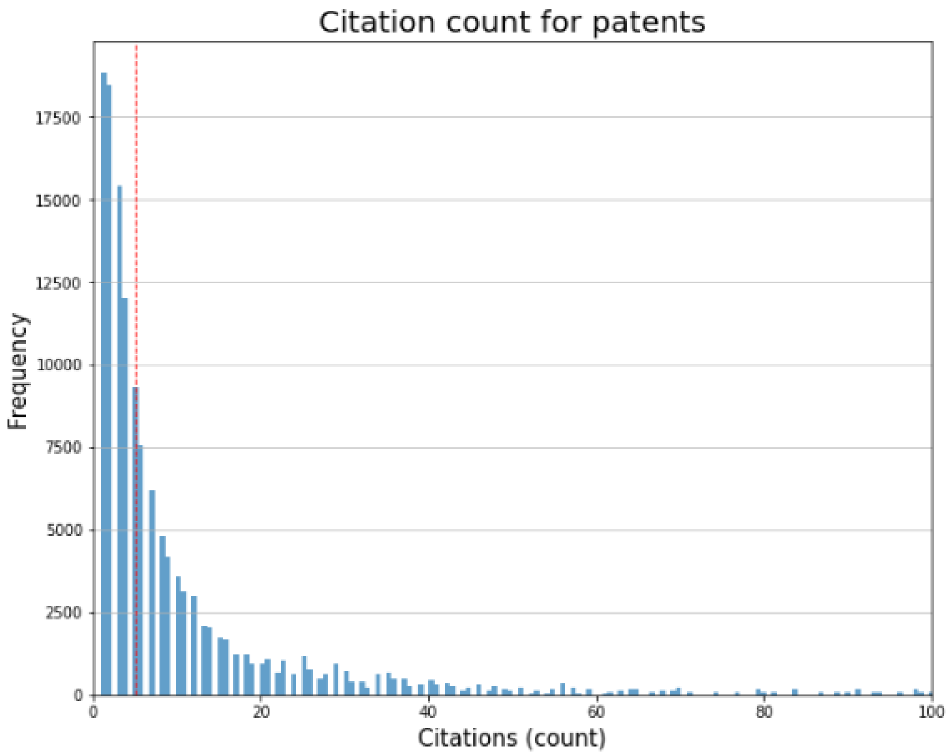


Figure 5 - The distribution of patent count for each cited patent in the dataset

Hypothesis 1 can then be considered through Column 3 as all three administrative boundaries are controlled for simultaneously. Here, the results show that a citation is much more likely from a patent within the local NUTS3 boundary than outside, as this is associated with a 26.7 percentage point increase in the probability of citation when controlling for both regional and country level effects. Therefore, this accounts for a greater than 50% increase in the baseline probability of citation of 0.500. For regional and country level boundaries, these results are 12.2 (~25%) and 5.4 (~10%) percentage points increases respectively, suggesting that knowledge spillovers are positively constrained by these boundaries. Therefore, given that all these boundaries are significant at the 1% significance level, this supports hypothesis 1 in suggesting that being in the same administrative boundary will have a significant and positive effect on knowledge spillovers in the UK. However, this effect is likely to be overstated because the effects of distance are not controlled for (Buzard, et al., 2020).

Column 4 is therefore used to show the full specification which includes the effects the distance, administrative boundaries and the 6-digit IPC match to explore both hypothesis 1 and 2 simultaneously. This shows that when controlling for both distance and boundary effects together, while they both remain significant, their strengths are smaller than when considered separately. Here, against the baseline of 0.500, being in the same local region as the cited patent results in an increase in the probability of citation by 21.7 (~40%) percentage points. For regional and country level matches these effects are 10.6 (~20%) and 3.6 (~7%) percentage points respectively, therefore supporting hypothesis 1. As for distance effects, moving from 0-25 miles to 25-50 miles results in a reduction of 8.8 (~18%) percentage point probability of citation. This decreases by a further 2.0 (10.8-8.8) percentage points moving from 25-50 to 50-100 miles. However, beyond this the effect of distance has no more depreciable effect suggesting that knowledge spillovers are concentrated within a boundary of 0-100 miles once administrative border effects are accounted for, supporting hypothesis 2.

It must also be noted that the effect of a match at the 6-digit IPC level remains positive and significant in these, and all subsequent, specifications. Therefore, this supports the idea that knowledge spills over more readily within the immediate technology class, but also that these results reflect actual knowledge spillovers rather than the existing technology distribution (Belenzon & Schankerman, 2013).

The final two columns build on this by exploring the suggestion that new and innovative pieces of knowledge or technology are less likely to be affected by distance because of their importance for firms and individuals (Carlino & Kerr, 2015). For this purpose, importance is measured through patent citation count, with a greater number of citations reflecting more important knowledge (Almeida & Kogut, 1999). For example, although the mean citation count for each patent is 3.34, figure 5 above shows that the number of citations is not evenly distributed. Therefore, using citation count as a measure of importance, the baseline specification is used to examine patents that are in the top 25% and bottom 25% for citations received in column 5 and 6 respectively.

Column 5 shows that although border effects are still positive and significant, distance is not important for highly valued knowledge. This can be seen in the fact that none of the distance boundaries are significant at the 5% significance level. Furthermore, the strength of the 6-digit IPC match is almost half of previous results, suggesting that more important knowledge is more likely to spillover to other firms in different industries. In contrast, patents that are in the bottom 25% for patent citations show even stronger distance decay than the baseline model, and that the strength of the 6-digit IPC match is stronger. This suggests that less important knowledge, as measured by patent citations, is unlikely to travel far or spillover to other industries, while more important knowledge, while limited by border effects, is unaffected by distance. Hence, this supports the suggestion that highly relevant and important knowledge will diffuse more widely in comparison to less important knowledge (Belenzon & Schankerman, 2013).

5.3) The effects of technology on knowledge spillovers

Another consideration is how the effect of distance on knowledge spillovers has evolved due to changes in technology (Charlot & Duranton, 2006). For this, figure 6 below shows that the dataset can be broken down into roughly five equal periods in terms of the application year of cited patents. Here, the final period stops at 2012 given the fact that the median lag time between the original and citing patent is 6 years, as can be seen in figure 7. Therefore, the application year for the cited patents is restricted to 2012 to allow the final originating year time to be cited, otherwise the results could be biased due to a reduced number of citations for the final year (Helmers, 2019). Thus, the role of technology on knowledge spillovers is explored in table 4 which uses the baseline model across the different years.

The first two columns explore this in terms of patents originating before and after 1994 respectively because 1994 is the median application year for cited patents prior to 2012. What these results suggest is that from 1994 onwards, knowledge spillovers decay much faster over distance than they have done previously, in contrast to hypothesis 3. This can clearly be seen in the first three distance

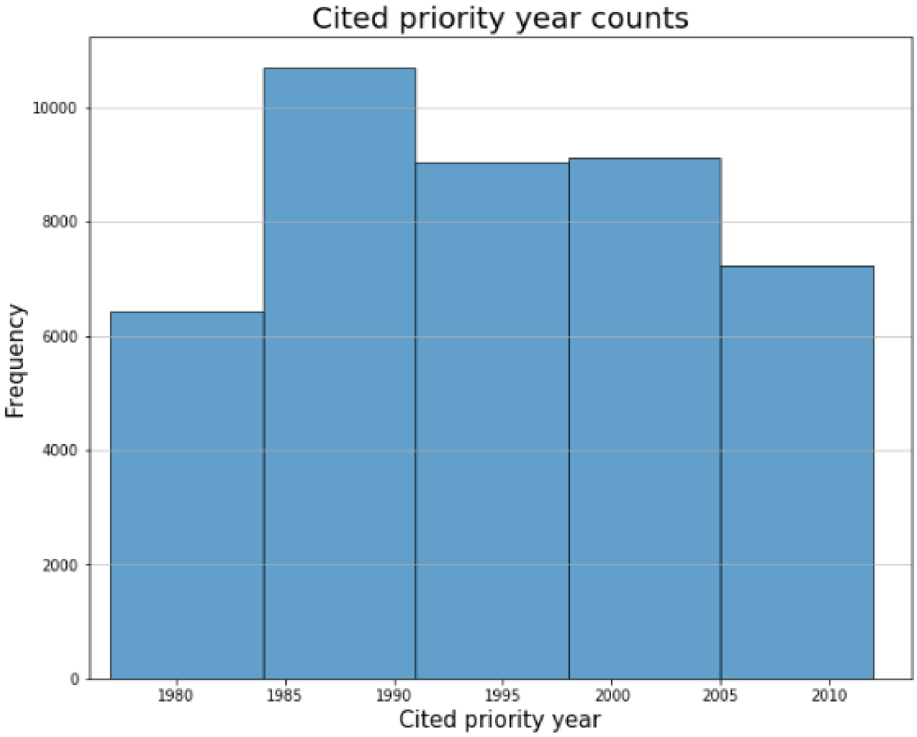


Figure 6 - A histogram showing the breakdown of cited patents into periods of 1977-1984, 84-91, 91-98, 98-2005, 05-12.

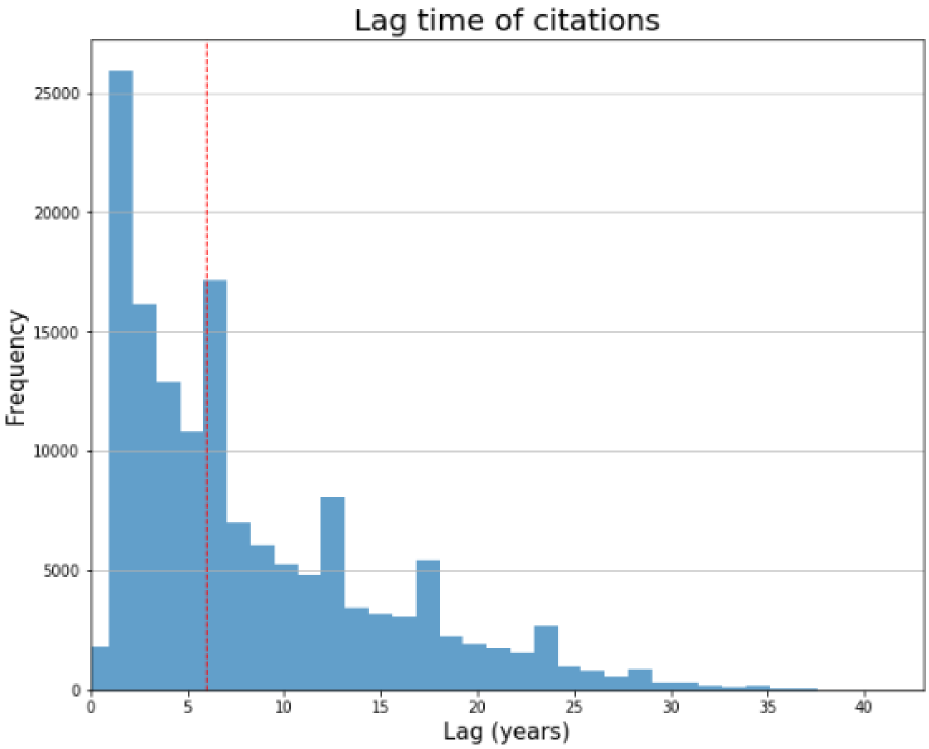


Figure 7 - A histogram showing the lag time in years between originating patents and their citation, with a mean of 6 years.

Table 4 – The effects of distance on citation probability across different periods

	Original patents prior to 1994 (1)	**Original patents after 1994 (2)**	**Original patents priority year 1977-84 (3)**	**Original patents priority year 1984-91 (4)**	**Original patents priority year 1991-98 (5)**	**Original patents priority year 1998-2005 (6)**	**Original patents priority year 2005-12 (7)**
Boundary match							
Local	0.2161	0.2183	**0.1388**	0.24	0.2518	0.1876	**0.3144**
	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***
Region	0.0925	0.1180	0.0951	0.1028	0.1066	0.1181	0.0901
	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.007)***
Country	0.0423	0.0269	**0.0525**	0.0384	0.0438	0.0096	**0.0189**
	(0.002)***	(0.002)***	(0.000)***	(0.001)***	(0.000)***	(0.465)	(0.319)
Distance (miles)							
25-50	**-0.0610**	**-0.1293**	**-0.1145**	-0.0226	-0.1168	-0.097	**-0.1556**
	(0.000)***	(0.000)***	(0.001)***	(0.447)	(0.001)***	(0.017)**	(0.005)***
50-100	**-0.0612**	**-0.1759**	**-0.0942**	-0.0501	-0.1259	-0.1542	**-0.2071**
	(0.000)***	(0.000)***	(0.003)***	(0.079)*	(0.000)***	(0.000)***	(0.000)***
100-250	**-0.0650**	**-0.1348**	**-0.1003**	-0.04	-0.104	-0.1407	**-0.1376**
	(0.000)***	(0.000)***	(0.001)***	(0.140)	(0.000)***	(0.000)***	(0.004)***
250-500	-0.0952	-0.1511	-0.1302	-0.0664	-0.124	-0.1572	-0.163
	(0.000)***	(0.000)***	(0.000)***	(0.016)**	(0.000)***	(0.000)***	(0.001)***
500-1000	-0.1050	-0.1470	-0.1456	-0.0735	-0.1289	-0.1544	-0.1442
	(0.000)***	(0.000)***	(0.000)***	(0.008)***	(0.000)***	(0.000)***	(0.003)***
1000-2500	-0.0876	-0.1273	-0.1514	-0.0513	-0.112	-0.1165	-0.1538
	(0.000)***	(0.000)***	(0.000)***	(0.069)*	(0.001)***	(0.000)***	(0.002)***
2500-5000	-0.1098	-0.1426	-0.1616	-0.0798	-0.1183	-0.1532	-0.1419
	(0.000)***	(0.000)***	(0.000)***	(0.004)***	(0.000)***	(0.000)***	(0.003)***
5000+	-0.1144	-0.1516	-0.1509	-0.0899	-0.1329	-0.1605	-0.1475
	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.002)***
Technology match							
IPC 6 match	**0.1434**	**0.1057**	**0.1713**	0.1381	0.106	0.1092	**0.097**
	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***
R^2	0.025	0.011	0.034	0.024	0.013	0.011	0.012
log likelihood	-90442	-104850	-27978	-48157	-50924	-43150	-24992
LLR p-value	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***
Observations	133784	153004	41790	71150	74412	62930	36506

*This table reports the coefficient estimates for the margin effects of a Probit regression model of citations to patents originating in the UK. This focuses on the results using the baseline specification across periods of 1) prior to 1993, 2) after 1993, 3) 1977-984, 4) 1984-1991, 5) 1991-1998, 6) 1998-2005, 7) 2005-2012. P-values are reported underneath the coefficients with statistical significance reported as * 10% , ** 5% and *** 1% levels.*

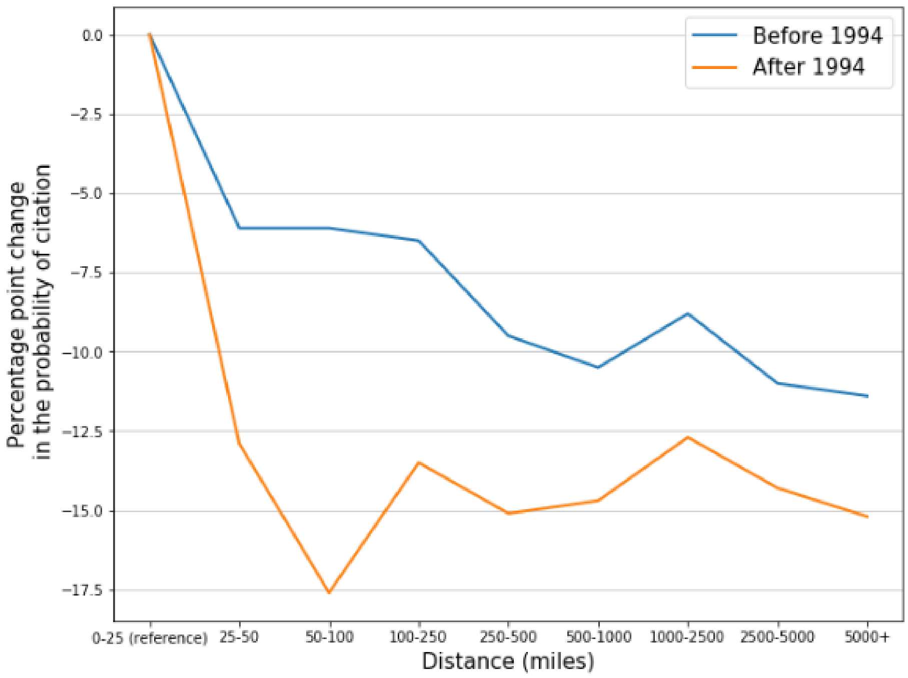


Figure 8 - This line graph shows the difference in percentage loss relative to the baseline for original patents with priority year before and after 1994.

boundaries where the effects are stronger by 6.8 (12.9-6.1), 11.5 (17.6-6.1) and 7.0 (13.5-6.5) percentage points respectively, as emphasised in figure 8. Furthermore, the importance of the 6-digit IPC match has decreased by 3.7 (14.3-10.6) percentage points between these periods suggesting that knowledge is more likely to cross over between industries than it did previously.

These results can be further specified by considering the five different periods from 1977 to 2012 as shown in columns 3 to 7 where, apart from column 4, the effects of distance and the effect of being in the same local area, increases gradually over each period. This therefore suggests that the importance of distance in knowledge spillovers has been increasing over time, supporting the conclusion of the first two columns and in contrast to hypothesis 3. However, the strength of the country effects and the match of the 6-digit IPC decrease over this period, suggesting that improvements in communication technologies have allowed for knowledge to flow across country and technology boundaries more easily than before. Overall, these results contrast with hypothesis 3 which suggests that distance and administrative boundaries would become less important over time.

5.4) The effects of time on knowledge spillovers

In figure 7, it can also be seen that 50% of citations occur within 6 years, but that some can take as long as 40 years. Therefore, this information can be used to explore hypothesis 4 in terms of knowledge spillovers spreading over time, as in table 5 below. Here, the first two columns look at cited-citing pairs that have a below and above median lag time respectively. What these results suggest is that knowledge spillovers are less constrained by distance the longer time goes on from the original creation of that knowledge. As can be seen in figure 9 below, a longer lag shows a less steep reduction in the effects of distance than those with a much shorter lag time, while the decay for both is mostly exhausted within 500 miles. Therefore, this supports the hypothesis that the longer the lag time between a patent and its citation, the less likely distance or administrative boundaries are to play a role in its dissemination.

This can be explored in more detail by breaking this down further into quartiles of lag time as seen in columns 3 to 6. These includes lag times of less than 3 years; between 3 and 6 years; between 6 and 12 years; and greater than 12 years respectively. This shows that until the lag time exceeds 12 years, distance is significant in explaining the probability of citation and that knowledge spillovers continue to decay up until the 500-mile boundary. However, the fourth column shows that by the time it comes to a lag time of 12 years or more distance is no longer significant in affecting the probability of citation. This therefore suggests that knowledge does disseminate further geographically over time, but that time must exceed 12 years for distance to no longer matter. However, it must be noted that the effect of administrative boundaries and the 6-digit IPC match strength remain relatively the same throughout all four time periods. This suggests that administrative boundaries remain important to knowledge access even if distance no longer does. Furthermore, if a technology does not cross industry boundaries initially, its likelihood of doing so in the future does not change significantly.

Table 5 – The effects of citation lag on patent citation

	Citation lag less than 6 years (1)	Citation lag longer than 6 years (2)	Bottom 25% for citation lag (3)	25-50% for citation lag (4)	50-75% for citation lag (5)	Top 25% for citation lag (6)
Boundary match						
Local	0.2371	0.1814	0.2452	0.2272	0.156	0.2098
	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***
Region	0.0984	0.1181	0.1056	0.088	0.1215	0.1135
	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***
Country	0.0461	0.0245	0.0606	0.0273	0.0088	0.0429
	(0.000)***	(0.005)***	(0.000)***	(0.024)**	(0.459)	(0.001)***
Distance (miles)						
25-50	-0.0959	-0.0734	-0.093	-0.0986	-0.108	-0.0275
	(0.000)***	(0.005)***	(0.001)***	(0.003)***	(0.002)***	(0.503)
50-100	-0.1137	-0.0959	-0.1157	-0.1086	-0.1459	-0.029
	(0.000)***	(0.000)***	(0.000)***	(0.001)***	(0.000)***	(0.445)
100-250	-0.1022	-0.0765	-0.1033	-0.0994	-0.1091	-0.0354
	(0.000)***	(0.001)***	(0.000)***	(0.001)***	(0.000)***	(0.328)
250-500	-0.1259	-0.0999	-0.1149	-0.1386	-0.1544	-0.0321
	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.376)
500-1000	-0.1282	-0.1014	-0.1175	-0.1412	-0.1586	-0.0313
	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.389)
1000-2500	-0.0992	-0.0911	-0.0894	-0.1107	-0.14	-0.029
	(0.000)***	(0.000)***	(0.001)***	(0.000)***	(0.000)***	(0.430)
2500-5000	-0.1108	-0.1189	-0.088	-0.1394	-0.193	-0.0558
	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.123)
5000+	-0.1208	-0.1172	-0.0991	-0.1478	-0.1643	-0.0574
	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.113)
Technology match						
IPC 6 match	0.1151	0.1323	0.1077	0.1237	0.1347	0.1295
	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***
R^2	0.0187	0.0147	0.0195	0.0182	0.0162	0.0135
log likelihood	-104520	-93970	-59744	-44756	-48095	-45863
LLR p-value	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***
Observations	153670	137596	87902	64768	70526	67070

*This table reports the coefficient estimates for the margin effects of a probit regression model of citations to patents originating in the UK, when the citation lag is 1) less than 6 years, 2) greater than 6 years, 3) less than 3 years (25%), 4) less than 6 years but greater than 3 (25-50%), 5) less than 12 years but greater than 6 years (50-75%), 6) greater than 6 years (75-100%). P-values are reported underneath the coefficients with statistical significance reported as * 10%, ** 5% and *** 1% level.*

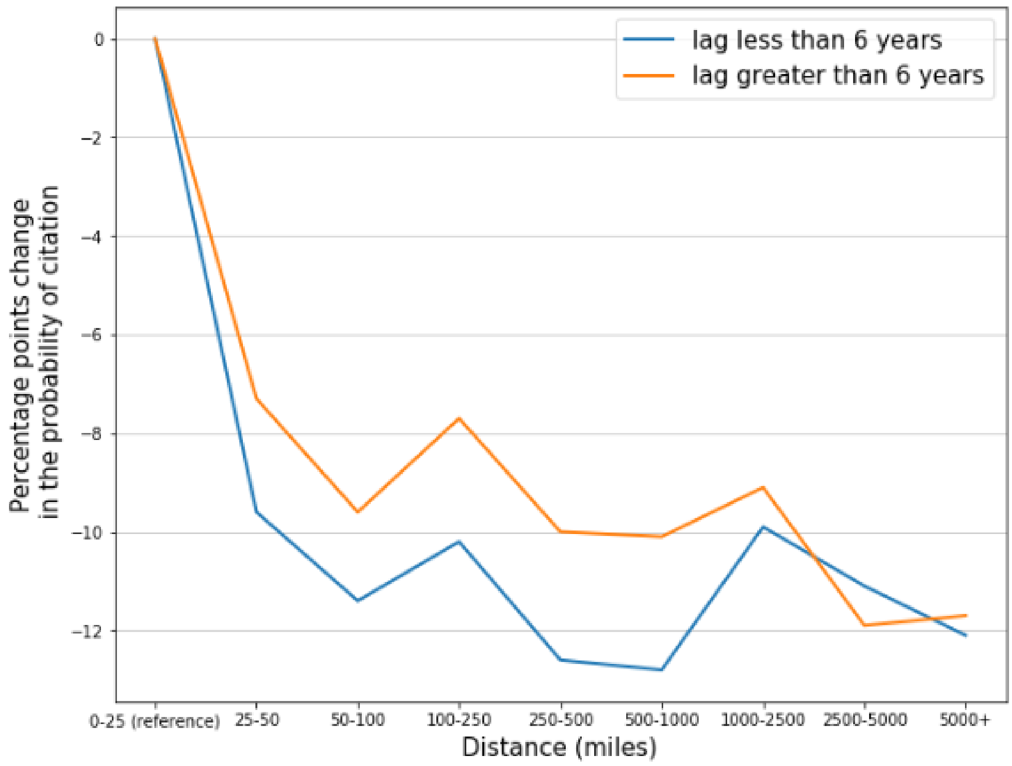


Figure 9 - The percentage points decrease in probability with a lag time of less than 6 years, against a lag time of greater than 6 years.

5.5) The effects of distance on knowledge spillovers in different industries

Finally, it is important to explore the different effects that distance and boundary effects may have across different industries as suggested in hypothesis 5. The results in table 6 below clearly support this hypothesis. For example, while in biotechnology administrative boundaries appear strongly positive and significant, distance has no significant effect on how knowledge may be transferred through knowledge spillovers. Furthermore, the match along the 6-digit IPC code is weak suggesting that knowledge can flow between different technologies within Biotechnology. Similar results can also be seen for the Medical industry.

However, in the case of the Pharmaceutical industry, distance does not appear to significantly negatively affect knowledge transmission until greater than 100 miles where in which there is a significance decrease of 11.1 percentage points. Furthermore, Engineering, Chemistry, IT, Telecommunications and Electrical industries show significant and strong effects of distance, along with administrative boundaries, although the strength of this effect differs as can be seen in figure

10 below. There are also differences in the strength of the 6-digit IPC matching effect across industries suggesting that in some industries, the need to match across similar technologies is more important than others. Overall, this suggests that the effect of distance on knowledge spillovers, as proxied by patent citations, differ across industries and this can have important implications for how policy may support different industries when encouraging clusters of firms.

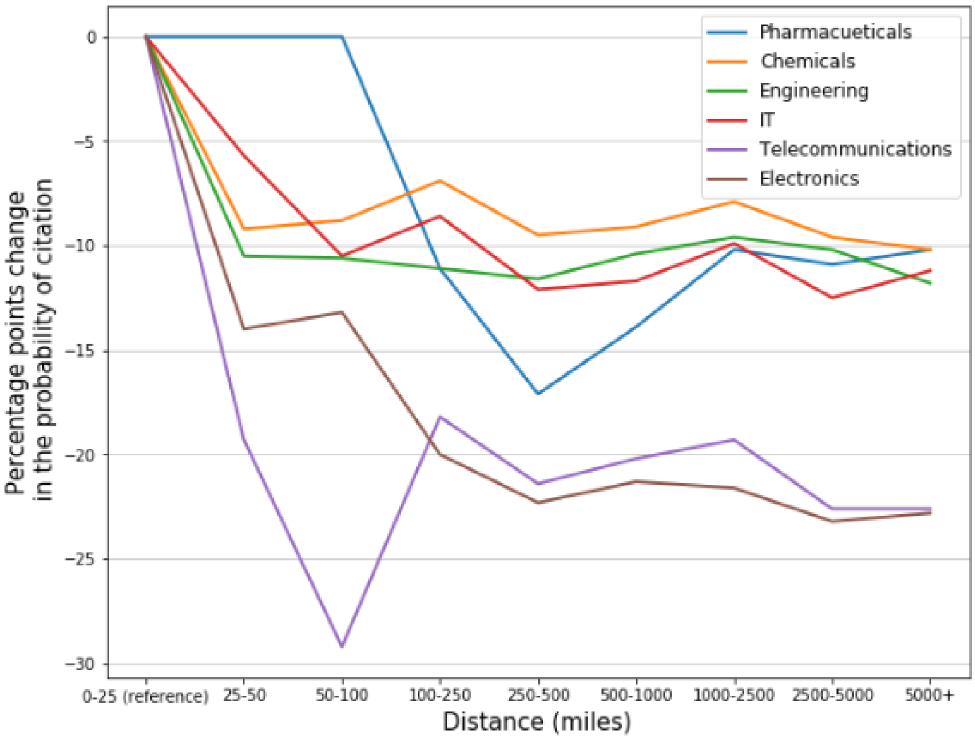


Figure 10 - The percentage point decrease in the probability of citation in the different industries where distance is significant, against the zero condition of 0-25 miles.

Table 6 –The effects of geography across industries

	Biotechnology	Medical	Pharmaceuticals	Chemicals (exc. Pharmaceuticals)	Engineering	IT	Telecommunications	Electronics
Boundary match								
Local	0.3059	0.2987	0.3268	0.1699	0.2056	0.246	-0.0488	0.2294
	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.680)	(0.000)***
Region	0.1629	0.0954	0.071	0.1651	0.1148	0.0531	0.0762	0.0412
	(0.001)***	(0.020)**	(0.043)**	(0.000)***	(0.000)***	(0.060)*	(0.202)	(0.155)
Country	0.0594	0.0339	0.0279	0.0417	0.0675	0.0292	0.0909	0.0257
	(0.033)**	(0.169)	(0.166)	(0.000)***	(0.000)***	(0.063)*	(0.006)***	(0.116)
Distance (miles)								
25-50	-0.0345	-0.0458	-0.0641	-0.0922	-0.1051	-0.0573	-0.1927	-0.14
	(0.738)	(0.534)	(0.270)	(0.004)***	(0.007)***	(0.231)	(0.103)	(0.005)***
50-100	-0.1137	-0.0198	-0.0896	-0.0881	-0.1061	-0.1053	-0.2919	-0.1315
	(0.243)	(0.777)	(0.120)	(0.003)***	(0.003)***	(0.022)**	(0.011)**	(0.006)***
100-250	-0.0587	-0.0356	-0.1107	-0.0686	-0.1108	-0.0865	-0.182	-0.2003
	(0.532)	(0.594)	(0.041)**	(0.017)**	(0.001)***	(0.049)**	(0.094)*	(0.000)***
250-500	-0.0361	-0.036	-0.1713	-0.0954	-0.1157	-0.1207	-0.2141	-0.2229
	(0.702)	(0.599)	(0.002)***	(0.001)***	(0.001)***	(0.006)***	(0.050)*	(0.000)***
500-1000	-0.0372	-0.0443	-0.139	-0.0906	-0.1035	-0.1169	-0.2022	-0.2126
	(0.693)	(0.518)	(0.011)**	(0.002)***	(0.003)***	(0.008)***	(0.064)*	(0.000)***
1000-2500	0.0091	-0.0305	-0.1021	-0.0793	-0.0956	-0.0985	-0.1931	-0.2163
	(0.924)	(0.658)	(0.065)*	(0.007)***	(0.007)***	(0.027)**	(0.080)*	(0.000)***
2500-5000	0.009	-0.0445	-0.1091	-0.0958	-0.1017	-0.1248	-0.2257	-0.2316
	(0.924)	(0.514)	(0.044)**	(0.001)***	(0.003)***	(0.004)***	(0.038)**	(0.000)***
5000+	-0.0237	-0.0412	-0.1017	-0.1026	-0.118	-0.1122	-0.2264	-0.2283
	(0.802)	(0.546)	(0.061)*	(0.000)***	(0.001)***	(0.010)**	(0.037)**	(0.000)***
Technology match								
IPC 6 match	0.0628	0.1578	0.0843	0.1291	0.1775	0.0857	0.0671	0.1017
	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.009)***
R^2	0.010	0.018	0.019	0.016	0.027	0.009	0.006	0.013
log likelihood	-7728.2	-12291	-12707	-51872	-35380	-28805	-6649.3	-27684
LLR p-value	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***
Observations	11264	18052	18696	76032	52464	41950	9654	40435

*This table reports the coefficient estimates for the margin effects of a probit regression model of citations to patents originating in the UK across different industries. P-values are reported underneath the coefficients with statistical significance reported as * 10%, ** 5% and *** 1% level. For patent industry classification, see Appendix D.*

6) Implications and discussion

These results above point to the importance that both distance and administrative boundaries have in terms of knowledge spillovers in the UK in relation to the five hypotheses developed. Therefore, discussions from these results can lead to four key insights that are important for UK innovation policy going forward. Firstly, in all complete models and regressions, local and regional administrative boundaries have significant and positive effects on the probability that a citation will occur, in line with hypothesis 1. This therefore supports previous research that comment on the importance of local, regional and country boundaries in that they may restrict knowledge flows within them (Buzard, et al., 2020). This is so even when controlling for the effects that distance may have on knowledge flows. Hence, it is important to understand how these administrative boundaries affect the results.

The focus on administrative boundaries and the role that they may play, while originally used to proxy for distance, has developed from evolutionary economics which emphasises the effects that institutions may have on economic growth and innovation (McCann & Van Oort, 2019). In this sense they suggest that such institutions and norms that affect knowledge spillovers can be constrained within these boundaries (Singh & Marx, 2013). Thus different localities and regions, even in the same region or country, can develop different institutional environments that would naturally restrict or support knowledge spillovers within their borders (Maskell, 2001). For example, the seminal work by Saxenian (1994) suggested that one of the differences between Route 128 and Silicon Valley in leading the latter to success was the atmosphere of collaboration and openness that was fostered by existing local institutions. This is further supported by work that looks at how knowledge spillovers may be influenced by worker mobility, where in which Silicon Valley also exhibited high levels of inter-firm mobility, therefore allowing for frequent changes by workers, bringing their knowledge and contacts with them (Breschi & Lissoni, 2000). This is significant because, similar to firm spin-outs which can transfer knowledge from one entity to another, these movement are often restricted by administrative boundaries regardless of distance, as they still want to be a part of existing institutions and norms that exist within city or region (Keeble & Wilkinson, 1999). Therefore, these results support the theory that local, regional and country level boundaries can influence knowledge spillovers because of the different institutions and norms that they may contain.

In terms of practical applications of this, policy-makers and academics must pay attention to how these institutions may differ across and within local areas and regions in the UK. For example, Saxenian (1989) examined the case of the 'silicon fen' within Cambridge, in which local spin-outs, labour mobility and knowledge transfers were positively supported by local institutions. Hence, policy must seek to learn from successful local institutional models that may facilitate knowledge transfer within the local area, to foster innovativeness and hence reap the economic rewards (Bathelt, 2005). In doing so, they must also pay attention to the scale at which policy aimed towards fostering innovation is applied either at the local, regional or country level, and hence which institutions operate at that scale (McCann & Van Oort, 2019). To this extent, the evidence here suggests that the main scale at which knowledge spillovers operate is the local level, to which policy could aim to support interactions here and more widely at the regional level. Furthermore, such policies must also acknowledge how cognitive, organisational and social proximity relate to institutional and geographical proximity in relation to knowledge spillovers (Boschma, 2005). Therefore, acknowledging the role that administrative boundaries may play in fostering institutions may allow policy to support within and between administrative boundary interaction to encourage innovative development and hence economic growth.

Secondly, in some results knowledge spillovers were seen to extend up to the 500-mile boundary, exceeding the expected distance in hypothesis 2. Indeed, examples from the United States suggest that knowledge spillovers would traditionally be expended within the 100-250 mile range (Belenzon & Schankerman, 2013), with some suggesting ranges of 0 to 25 miles (Buzard, et al., 2020) and others within a much smaller range of even 750 meters (Arzaghi & Henderson, 2008). It is therefore surprising to find that in the case of the UK, which is much smaller than the US, knowledge spillovers do not decay fully until a much greater distance. However, while some results show decay extending up to this distance, most of this decay is exhausted between the 25-50 or 50-100-mile boundary, in line with hypothesis 2. This would therefore support the importance of face-to-face interaction as these distances can be seen to represent an extended commuting boundary (Belenzon & Schankerman, 2013), or at the least a range in which a business meeting or event can occur within a single day. Hence, within this range, repeated face-to-face interaction can occur which is suggested to be the key factor driving the transfer of tacit knowledge that leads to knowledge spillovers (Venables & Storper, 2004). Thus, distance is seen to play a significant part in knowledge spillovers.

However, the range seen in some results may suggest the influence that both extended networks and temporary colocation may have on how knowledge is exchanged. This is because repeated face-to-face interaction cannot always occur at such distances and therefore other events or modes of communication may be present in influencing these results. For example, significant research has begun to explore the role that long-distance networks may have on knowledge access from different clusters through 'pipelines' (Bathelt, et al., 2004; Speldekamp, et al., 2020). This refers to channels that may be used in distant interactions to gain access to knowledge that is emerging in areas that are not geographically proximate, potentially due to other forms of proximity, such as those described by Boschma (2005) as mentioned above. This can therefore reduce the negative costs associated with interaction across significant distances and hence allow for knowledge spillovers to occur at a much greater scale (Boschma & Ter Wal, 2007). These relationships could also be reinforced through temporary agglomerations such as hackathons, trade fairs and industry events that temporarily bring together typically geographically distant individuals but are proximate in these other ways (Breschi & Lissoni, 2003). These events work by establishing common local codes and languages that temporarily reduce the cost of transaction and communication, reducing the costs associated with developing extra-local pipelines (Henn & Bathelt, 2015). Therefore, this could potentially explain some of the large distances that are seen, but that are within the potential for visits to short events or meetings.

Hence, policy-makers must acknowledge the potential that networks and temporary agglomerations may play in knowledge transfer and spillovers. Policy may support knowledge spillovers across these greater distances by providing resources and support that would allow firms to develop connections beyond their extended commuting boundary. This is important because, as Bathelt et al., (2004) note, there are often significant costs associated with developing global pipelines, especially for smaller firms (Maskell, et al., 2006). The recognition of this need can already be seen in the development of the UK catapult network (Catapult network, 2017). These aid the development of networks for industries that the UK is considered world leaders in, as they are designed to foster connections between universities, research institutions, businesses and public bodies with the aim of supporting productive collaboration. To this end, they have already being accredited with creating and supporting 2,473 industry collaborations in the UK (Catapult network, 2017), with early signs of success from government created Knowledge Performance Indicators (Bailey & Tomlinson, 2017). However, it has been noted that further work can be done in terms of engaging with SMEs and providing support for attendance at larger gatherings (Bailey & Tomlinson, 2017). This could include the continued development of the Catapult network to help foster pipelines between institutions, the provision of innovation managers within clusters whose job it is to find and support connections between firms similar to the role seen in an incubator model (McCann, 2013), and support for SMEs to attend extra-local events.

Furthermore, in contrast to hypothesis 3, these results show that the effects of distance and boundary conditions have become stronger suggesting that technology is not leading to ‘the death of distance’ but instead reinforcing its effects (Charlot & Duranton, 2006). This result is contrary to other findings, such as in Keller (2002) where they suggest that localisation effects of technology diffusion has declined over their period of study from 1970 to 1995, and Thompson (2006) who provide evidence that the strength of localisation in the US has decreased over time. Thus, the argument that the cost of exchanging knowledge and goods has decreased such that knowledge spillovers would be less impeded by distance (Howells, 2002). Therefore, the finding here that distance and boundary conditions are becoming stronger over time is surprising.

One theory that could potentially explain these results is the way in which new communication methods and media are used. This is because, as Charlot and Duranton (2006) suggest, instead of technology allowing for communication across longer distances to transfer knowledge, it can be used to reinforce existing channels of communication made through face-to-face interaction. Thus, they also suggest face-to-face communication is unlikely to be replaced by these devices for interactions that would transfer complex information. Hence, instead of technology leading to the 'death of distance', it could be reinforcing the effects that distance has on already existing communication channels. This is especially so where although new technologies allow greater access to information at further distances there are still significant costs associated with learning and integrating that knowledge (Bathelt, et al., 2004).

Furthermore, these results are also in line with findings by Sonn and Storper (2008) who show that in the US, the importance of proximity at the Country, State and Metropolitan Statistical Area levels has increased over time. In doing so they attribute this result to the shortening of technology life cycle. This is because of the result that is supported in many other papers, and weakly supported here, that the shorter the lag between the original and citing patent, the more local it is likely to be. For example, in the case of the electronics industry where there is a short half-life for new technologies, evidence suggests that knowledge spillovers decay much faster than in industries where this is not the case, such as in Pharmacology (Jaffe & Trajtenberg, 1996). Therefore, if technology life cycles are shortening, quick access to new information and technologies is likely to become more important, and hence distance also.

The policy implications for this are similar to those regarding network and temporary agglomeration solutions above. This is because if technology is reinforcing existing communication networks that are the result of repeated face to face interaction, then policies such as supporting temporary agglomerations and providing resources to help foster pipelines between firms and agglomerations could extend the geographic distance over which they operate. This is especially so if the conclusions drawn by Sonn and Storper (2008) and Jaffe and Trajtenberg (1996) hold, such that technology lifecycles are shortening, and that this is associated with more localised knowledge spillovers. Therefore, for these industries and technologies, solutions must be created that would support quick and rapid diffusion of new knowledge and ideas so that the benefits of these can be spread more widely. Alternatively, support can be provided for industries that exhibit these characteristics to allow them to collocate in areas, such as seen for Biotechnology in Cambridge (Deloitte, 2015). The emphasis therefore is on allowing knowledge to diffuse between firms and individuals at a much faster pace through either supporting colocation or developing institutions that can support the geographical extension of individual's networks.

This leads into the final implication of the differences in importance and strength of boundaries and distance in knowledge transfer across different industries (Duranton & Overman, 2005; McCann & Van Oort, 2019). This result could relate to the suggestion above of the different technological lifecycles affecting distance or the fact that the importance of patenting and therefore whether knowledge spillovers can be represented by patents will be affected by the nature of the industry (Jaffe, et al., 2000). Furthermore, industries that depend more on codification such as Pharmaceuticals or medical technologies may be those in which knowledge spillovers are less affected by distances (Howells, 2002). Therefore, these results point to the fact that knowledge spillovers within different industries are likely to be influenced by their nature, and that as such policy must pay attention to these nuances for it to be effective. For example, given the importance of the local boundary conditions in Biotechnology, emphasis could be placed on local institutions that foster interaction, such as within the Cambridge locality (While, et al., 2004). However, in order to reach conclusive results, future work may seek to use alternative methods and measures to examine this in more detail across multiple different industries at different scales.

Finally, it is also important to acknowledge existing circumstances due to Covid-19 in relation to this study. This is because the findings here suggest that face-to-face interaction is important, and that current evidence suggests technology is unlikely to replace this. Therefore, with face-to-face interaction limited by current circumstances, leading to an increase in work-from-home and the use of technology to communicate, based on these results there may be a reduction in innovative output. This is because of a potential fall of the cross-fertilisation of ideas that may have otherwise occurred due to serendipitous transfers of knowledge through face-to-face interaction in daily life (Catalini, 2018). Hence, this situation must be monitored going forward, and may present itself to an interesting research study to be conducted in the future in terms of whether technology can truly become a working substitute for face-to-face interaction.

7) Conclusion

Existing work has shown that knowledge spillovers play an important role in supporting innovation and economic growth at the local, regional and country level. This study contributes to this literature by utilising existing methods and extending them by using a Probit regression and applying this to a case study of the UK. What these results show is that distance and administrative boundaries are significant determinants of knowledge spillovers within the UK, to the extent that most knowledge spillovers are exhausted within 100 miles, and that local, regional and country level boundaries are significant in most specifications. This includes the fact that effects of distance have become stronger since 1977, with knowledge spillovers decaying faster in more recent years. Furthermore, distance is significant until 12 years after the original patent was published, supporting the theory of the diffusion of knowledge. Finally, the effects of distance and administrative boundaries can vary by industry, with some industries exhibiting strong evidence of localisation of knowledge spillovers when compared to others.

These results mean that policy must pay attention to the effects that geography may have on innovation outcomes. This includes the acknowledgement that institutional factors, that are seen to be constrained by administrative boundaries within the literature, can influence the degree to which knowledge is likely to be transferred between firms. As such, policy must work within existing institutional constraints at the level at which policy is applied and seek to foster the development of positive institutional factors within other boundaries, or bridge the gap in terms of cognitive, organisational, social or institutional proximity between firms. Furthermore, given the restriction of most knowledge spillovers within an extended commuting boundary, face-to-face interaction is important for transferring tacit knowledge. Thus, to extend the geographical extent of diffusion, policy can include the support for temporary agglomerations, whilst also providing resources to support the development knowledge pipelines. Finally, policy must also pay attention to the industry to which these policies are most likely to affect, with different industries exhibiting different effects of geography. Therefore, solutions must adapt to suit the culture and institutional background of these industries.

It is worth noting, however, that this study has several limitations that could allow for future research. Firstly, in exploring the effects of knowledge spillovers in the UK there was no geographical restriction placed on citations. While industrial distribution was controlled for using the control dataset, comparison with the results from studies covering the US, in terms of the percentage of citations originating within the country, suggests that future research could explore the results when constricting the data to patents that originate in Europe. This could therefore also extend into an exploration of the determinants of knowledge spillovers within the wider European context which could seek to inform the EU's policy of continued economic integration and their innovation policy (Barca, 2009).

Furthermore, this research focuses on the effects of geographical proximity in knowledge spillovers, whereas previous research suggests that factors such as cognitive, organisational, social and institutional factors may also play a part (Boschma, 2005). Hence, future specifications could include variables in the regression to account for these factors, such as by including a dummy variable that would examine the relationship between local or regional specialisations as measured by location quotients, therefore being able to see how distance measure are subsequently affected by this (Moreno, et al., 2005).

Finally, the results of how distance and administrative boundaries affect knowledge spillovers are interpreted through existing theories. However, to date there has been little research that has explored the mechanisms and channels through which these factors influence knowledge spillovers directly (Singh & Marx, 2013). This could therefore include highly granular research that explores how knowledge may be transferred through different methods of communication to be able to understand how knowledge spillovers work. This would be able to inform policy more clearly as to what mechanisms of interaction are most efficient for knowledge transfer.

Bibliography

Adams, J. & Jaffe, A., 1996. Bounding the effects of R&D: An investigation using matched establishment-firm data. *National Bureau of economic research,* Volume No. w5544.

Agrawal, A., Cockburn, I., Galasso, A. & Oettl, A., 2014. Why are some regions more innovative than others? The role of small firms in the presence of large labs. *Journal of Urban Economics,* Volume 81, pp. 149-165.

Allison, P., 2017. *Statistical Horizons.* [Online]
Available at: https://statisticalhorizons.com/in-defense-of-logit-part-1
[Accessed 29 June 2020].

Almeida, P. & Kogut, B., 1999. Localisation of knowledge and the mobility of Engineers in Regional Networks. *Management Science,* 45(7), pp. 905-917.

Armstrong, H. & Taylor, J., 2000. *Regional Economics and Policy.* 3rd ed. Oxford: Blackwell.

Arzaghi, M. & Henderson, J., 2008. Networking off Madison Avenue. *Review of Economic Studies,* Volume 75, pp. 1011-1038.

Bailey, D. & Tomlinson, P., 2017. Back to the future? UK industrial policy after the great financial crisis. In: P. Arestis & M. Sawyer, eds. *Economic Policies since the Global Financial Crisis. International Papers in Political Economy.* Cham: Palgrave Macmillan, pp. 221-263.

Barca, F., 2009. *An Agenda for a reformed cohesion policy,* s.l.: s.n.

Bathelt, H., 2005. Geogrpahies of production: growth regimes in spatial perspective (II) - knowledge creation and growth in clusters. *Progress in Human Geography,* 29(2), pp. 204-216.

Bathelt, H., Malmberg, A. & Maskell, P., 2004. Clusters and Knowledge: local buzz, global pipelines and the process of knowledge creation. *Progress in Human Geography,* 28(1), pp. 31-56.

Belenzon, S. & Schankerman, M., 2013. Spreading the word: geography, policy and knowledge spillovers. *The Review of Economics and Statistics,* 95(3), pp. 884-903.

Bell, G., 2005. Clusters, networks, and firm innovativeness. *Strategic management journal,* Volume 26, pp. 287-295.

Boschma, R., 2005. Proximity and Innovation: A critical assessment. *Regional studies,* 39(1), pp. 61-74.

Boschma, R. & Ter Wal, A., 2007. Knowledge Networks and Innovative Performance in an Industrial District: The Case of a Footwear District in the South of Italy. *Industry and Innovation,* 14(2), pp. 177-199.

Breschi, S. & Lissoni, F., 2000. *Knowledge Spillovers and Local Innovation systems.* Barcelona, European Regional Science Association.

Breschi, S. & Lissoni, F., 2003. Mobility and Social Networks: Localised knowledge spillovers revisited. *CESPRI.*

Buzard, K. et al., 2017. The Agglomeration of American R&D labs. *Journal of Urban Economics,* Volume 101, pp. 14-26.

Buzard, K. et al., 2020. Localised knowledge spillovers: Evidence from the spatial clustering of R&D labs and patent citations. *Regional Science and Urban Economics,* Volume 81.

Cappelli, R., Czarnitzki, D. & Kraft, K., 2014. Sources of spillovers for imitation and innovation. *Research Policy,* Volume 43, pp. 115-120.

Carlino, G. & Kerr, W., 2015. Agglomeration and Innovation. In: G. Duranton, J. Henderson & W. Strange, eds. *HAndbook of Regional and Urban Economics.* s.l.:s.n., pp. 349-404.

Catalini, C., 2018. Microgeography and the Direction of Inventive Activity. *Management Science,* 64(9), pp. 4348-4346.

Catapult network, 2017. *Catapult.org.* [Online]
Available at: https://s3-eu-west-1.amazonaws.com/media.www.catapult/wp-content/uploads/2017/08/02164038/Cross-Catapult-Network-Executive-Summary-2017.pdf
[Accessed 19 February 2019].

Charlot, S. & Duranton, G., 2006. Cities and Workplace Communication: Some Quantitative French Evidence. *Urban Studies,* 43(8), pp. 1365-1394.

Cohen, W. & Levinthal, D., 1990. Absorptive capacity: A new Perspective on Learning and Innovation. *Administrtaive Science Quarterly,* 35(1), pp. 128-152.

Cortinovis, N. & Van Oort, F., 2018. Between spilling over and boiling down: network-mediated spillovers, local knowledge base and productivity in European region. *Journal of Economic Geography,* Volume 0, pp. 1-28.

Cotropia, C., Lemley, M. & Sampat, B., 2013. Do applicant patent citations matter?. *Research Policy,* Volume 42, pp. 844-854.

Cottineau, C. et al., 2019. Defining urban clusters to detect agglomeration economies. *Urban Analytics and City Science,* 46(9), pp. 1611-1626.

Deloitte, 2015. *Genomics in the UK: an industry study for the Office of Life Sciences,* London: Government.

Duguet, E. & MacGArvie, M., 2005. How well do patent citations measure flows of technology? Evidence from french innovation surveys. *Economics of Innovation and New Technology,* 14(5), pp. 375-393.

Duranton, G. & Overman, H., 2005. Testing for Localization Using Micor-Geographic Data. *Review of Economic Studies,* Volume 72, pp. 1077-1106.

Eisenberg, R., 2020. *Forbes.* [Online]
Available at: https://www.forbes.com/sites/nextavenue/2020/04/10/is-working-from-home-the-future-of-work/#16cc893c46b1
[Accessed 24 June 2020].

Eriksson, R. & Hane-Weijman, E., 2017. How do regional economies respond to crises? The geography of job creation and destruction in Sweden (1990-2010). *European and Regional Studies,* 24(1), pp. 87-103.

Eurostat, 2020. s.l.: s.n.

Feldman, M. & Kogler, D., 2010. Stylized Facts in the Geography of Innovation. In: H. Bronwyn, H. Hall & N. Rosenberg, eds. *Handbook of the Economics of Innovation.* s.l.:s.n., pp. 381-410.

Feser, E., 1998. Old and new theories of industry clusters. In: M. Steiner, ed. *Clusters and regional specialistion.* London: Pion, pp. 18-40.

Frenken, K., Van Oort, F. & Verburg, T., 2007. Related Variety, Unerlated Variety and Regional Economic Growth. *Regional Studies ,* 41(5), pp. 685-697.

Fritsch, M. & Franke, G., 2004. Innovation, regional knowledge spillovers and R&D cooperation. *Research Policy,* Volume 33, pp. 245-255.

GADM, 2020. s.l.: GADM.

Geroski, P., 2000. Models of technology diffusion. *Research Policy,* 29(4-5), pp. 603-625.

Glaeser, E., 1999. Learning in cities. *Journal of Urban Economics,* 46(2), pp. 254-277.

Glaeser, E., 2010. *Agglomeration Economics.* 1st ed. Chicago: The University of Chicago Press.

Glaeser, E. & Kohlhase, J., 2004. Cities, Regions and the Decline of Transport costs. *Papers in Regional Science,* Volume 83, pp. 197-228.

Glaeser, E. & Mare, D., 2001. Cities and Skills. *Journal of Labour Economics,* 19(2).

Gordon, I. & McCann, P., 2000. Industrial Clusters: Complexes, Agglomeration and/or Social Networks. *Urban Studies,* 37(3), pp. 513-532.

Graham, S., Merges, R., Samuelson, P. & Sichelman, T., 2009. High Technology Entrepreneurs and the patent system: results of the 2008 Berkeley PAtent Survey. *Berkley Technology Law Journal,* 24(4), pp. 255-327.

Hannigan, T., Cano-Kollmann, M. & Mudambi, R., 2015. Thriving innovation amidst manufacturing decline: the Detroit auto cluster and the resilience of local knowledge production. *Industrial and Corporate Change,* 24(3), pp. 613-634.

Harrison, B., 1992. Industrial districts: old wine in new bottles?. *Regional studies,* 26(5), pp. 469-483.

Helmers, C., 2019. Choose the Neighbor before the House: Agglomeration Externalities in a UK Science Park. *Journal of Economic Geography,* Volume 19, pp. 31-55.

Helsey, R. & Strange, W., 1990. Agglomeration economies and matching in a system of cities. *Reg Sci Urban Econ,* Volume 20, pp. 189-212.

Henderson, J., 2007. Understanding knowledge spillovers. *Regional Science and Urban Economics,* Volume 37, pp. 497-508.

Henn, S. & Bathelt, H., 2015. Knowledge generation and field reproduction in temporary clusters and the role of business conferences. *Geoforum,* Volume 58, pp. 104-113.

Hern, A., 2020. Covid-19 could cause permanent shift towards home working. *The Guardian*, 13 March.

Herstad, S., Aslesen, H. & Ebersberger, B., 2014. On industrial knowledge bases, commercial opportunities and global innovation network linkages. *Research Policy,* Volume 43, pp. 494-504.

Hettinger, E., 1989. Justifying intellectual property. *Philosphy & public affairs,* 18(1), pp. 31-52.

Howells, J., 2002. Tacit Knowledge, Innovation and Economic Geography. *Urban Studies,* 39(5-6), pp. 871-884.

Inaba, T. & Squicciarini, M., 2017. *ICT: A new taxonomy based on the international patent classification,* Paris: OECD Publishing.

Jacobs, J., 1961. *The death and life of great american cities.* 1st ed. New York: Random House.

Jaffe, A. & Trajtenberg, M., 1996. Flows of Knowledge from universities and federal laboratories: Modeling the flow of patent citations over time and across institutional and geographic boundaries. *Procedure of the National Acadamy of Science,* Volume 93, pp. 12671-12677.

Jaffe, A., Trajtenberg, M. & Fogarty, M., 2000. *THE MEANING OF PATENT CITATIONS: REPORT ON THE NBER/CASE-WESTERN RESERVE SURVEY of PATENTEES,* Cambridge: NBER.

Jaffe, A., Trajtenberg, M. & Henderson, R., 1993. Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *The Quarterly Journal of Economics,* 108(3), pp. 577-598.

Keeble, D. & Wilkinson, F., 1999. Collective learning and knowledge devlopment in the evolution of regional clusters of high technology SMEs in Europe. *Regional Studies,* pp. 295-303.

Keeble, D. & Wilkinson, F., 2000. High-Technology SMEs, Regional Clustering and Collective Learning: An Overview. In: D. Keeble & F. Wilkinson, eds. *High-Technology Clusters, Networking and Collective Learning in Europe.* Abingdon: Routledge, pp. 1-20.

Kerr, W. & Kominers, S., 2015. AGGLOMERATIVE FORCES AND CLUSERR SHAPES. *The Review of Economics and Statistics,* 97(4), pp. 877-899.

Kogler, D., 2015. Intellectual property and patents: knowledge creation and diffusion. In: J. Bryson, J. Clark & V. Vanchan, eds. *Handbook of Manufacturing Industries in the World Economy.* s.l.:Edward Elgar Publishing, pp. 163-191.

Krugman, P., 1998. What's new about the New Economic Geograhy. *Oxford Review of Economic Policy,* 14(2), pp. 7-17.

Leeper, T., 2018. *CRAN.* [Online]
Available at: https://cran.r-project.org/web/packages/margins/vignettes/TechnicalDetails.pdf
[Accessed 29 June 2020].

Mackinnon, D. & Cumbers, A., 2011. *Introduction to economic geography: globalization, uneven development and place.* Edinburgh: Pearson education limited.

Malmberg, A., 1996. Industrial geography: agglomeration and local milieu. *Progress in human geography,* 20(3), pp. 392-403.

Malmberg, A. & Maskell, P., 1997. Towards an explanation of regional specialisation and industry agglomeration. *European Planning Studies,* 5(1), pp. 25-42.

Malmberg, A. & Maskell, P., 2002. The elusive concept of localisation economies: towards a knowledge-based theory of spatial clustering. *Environment and Planning,* Volume 34, pp. 429-449.

Malmberg, A. & Power, D., 2005. How do firms in clusters create knowledge. *Industry and innovation,* 12(4), pp. 409-431.

Marshall, A., 1890. *Industial organisation continued.* 1st ed. London: Macmillan.

Martin, R. & Sunley, P., 2003. Deconstructing clusters: chaotic concept or policy panacea?. *Journal of Economic Geography,* 3(1), pp. 5-35.

Maskell, P., 2001. Towards a knowledge-based theory of the geographical cluster. *Industrial and Corporate Change,* 10(4), pp. 921-943.

Maskell, P., Bathelt, H. & Malmberg, A., 2006. Building global knowledge pipelines: the role of temporary clusters. *European Planning Studies,* 14(8), pp. 997-1013.

Maskell, P. & Malmberg, A., 1995. *Localised learning and industrial competitiveness, Working Paper 80,* Berkeley, CA: Berkeley Roundatble on the International Economy (BRIE).

McCann, B. & Folta, T., 2008. Location Matters: Where We Have Been and Where We Might Go in Agglomeration Research. *Journal of Managament,* 34(3), pp. 532-565.

McCann, P., 2013. *Modern Urban and Regional Economics.* 2nd ed. Oxford: Oxford University Press.

McCann, P. & Van Oort, F., 2019. Theories of agglomeration and regional eocnomic growth: a historical review. In: R. Capello & P. Nijkamp, eds. *Handbook of Regional Growth and Development theories.* Cheltenham: Edward Elgar Publishing Limited, pp. 6-23.

Miguelez, E. & Moreno, R., 2015. Knowledge flows and the absorptive capacity of regions. *Research policy,* Volume 44, pp. 833-848.

Moreno, R., Paci, R. & Usai, S., 2005. Spatial spillovers and innovation activity in European regions. *Environment and Planning A,* Volume 37, pp. 1793-1812.

Murata, Y., Nakajima, R., Okamoto, R. & Tamura, R., 2014. Localised knowledge spillovers and patent citations: a distance-based approach. *The Review of Economics and Statistics,* 96(5), pp. 967-985.

O'Clery, N., Heory, S., Hulot, F. & Beguerisse-Diaz, M., 2019. Unravelling the forces underlying urban industrial agglomeration.

OECD.stat, 2020. *OECD.stat.* [Online]
Available at: https://stats.oecd.org/Index.aspx?DataSetCode=PATS_IPC#
[Accessed 02 July 2020].

OECD, 2020. s.l.: s.n.

OECD, January 2020. *Citations database.* s.l.:s.n.

OECD, January 2020. *OECD REGPAT Database - background information.* s.l.:OECD.

OECD, January 2020. *REGPAT database.* s.l.:s.n.

O'Sullivan, A., 2012. *Urban Economics.* 8th ed. New York: McGRaw-Hill/Irwin.

Ottavino, G. & Puga, D., 2003. Agglomeration in the Global Economy: A Survey of the 'New Economic Geography'. *The World Economy,* 21(6).

Owen-Smith, J. & Powell, W., 2004. Knowledge Networks as Channels and Conduits: The effects of spillovers in the Boston Biotechnology Community. *Organisation Science,* 15(1), pp. 5-21.

Park, J. W. I. et al., 2019. Global labor flow network reveals the hierarchical organisation and dynamics of geo-industrial clusters. *Nature Communications,* Volume 10.

Parr, J., 2002. Agglomeration economies: ambiguities and confusions. *Environment and planning,* Volume 34, pp. 717-731.

Polanyi, M., 1958. *Personal Knowledge: Towards a Post-Critical Philosophy.* London: Routledge & Kegan Paul.

Porter, M., 1998. Clusters and the new economics of competition. *Harvard Business Review,* 76(6), pp. 77-90.

Porter, M., 2000. Locations, clusters, and company strategy. In: G. Clark, M. Feldman & M. Gertler, eds. *The Oxford Handbook of Economic Geography.* Oxford: Oxford University Press, pp. 253-274.

Posner, R., 2005. Inetllectual property: the law and economics approach. *Journal fo economic perspectives ,* pp. 57-73.

Puga, D., 2010. The magnitude and causes of agglomeration economies. *Journal of Regional Science,* 50(1), pp. 203-219.

Rammer, C., Kinne, J. & Blind, K., 2020. Knowledge proximity and firm innovation: A microgeographic analysis for Berlin. *Urban Studies,* 57(5), pp. 996-1014.

Regan, A. & Harby, J., 2020. *BBC.* [Online]
Available at: https://www.bbc.co.uk/news/uk-england-52854117
[Accessed 24 June 2020].

Roberts, M., 2004. The Growth Performance of teh GB Counties: Some New EMpirical Evidence fro 1977-1993. *Regional studies,* 38(2).

Rosenthal, S. & Strange, W., 2004. Evidence on the nature and source of agglomeration economies. *Handbook of regional and urban economics,* Volume 4, pp. 2119-2171.

Saxenian, A., 1989. The Chesire cat's grin: innovation, regional development and the cambridge case. *Economy and society,* 18(4), pp. 448-477.

Saxenian, A., 1994. *Regional advantage: Culture and competition in Silicon Valley and Route 128.* 1st ed. Cambridge, MA: Harvard University Press.

Singh, J. & Marx, M., 2013. Geographic Constraints on Knowledge Spillovers: Political Borders vs. Spatial Proximity. *Management Science,* 59(9), pp. 2056-2078.

Sonn, J. & Stoprer, M., 2008. The increasing importance of geographical proximity in knowledge production: an analysis of US patent citations, 1975-1997. *Environment and Planning A,* Volume 40, pp. 1020-1039.

Speldekamp, D., Saka-Helmhout, A. & Knoben, J., 2020. Reconciling Perspectives on Clusters: An Intergrative Review and Research Agenda. *International Journal of Management Reviews,* Volume 22, pp. 75-98.

Sweet, C. & Eterovic, D., 2019. Do patent rights matter? 40 years of innovation, complexity and productivity. *World Development,* Volume 115, pp. 78-93.

Talibova, R., Qu, B., Ding, J. & Lan, S., 2018. *Github.* [Online]
Available at: https://jbhender.github.io/Stats506/F18/GP/Group14.html
[Accessed 6 July 2020].

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

Thompson, P. & Fox-Kean, M., 2005. Patent Citations and the Geography of knowledge spillovers: A Reassessment. *The American Economic Review,* 95(1), pp. 450-460.

Tyler, P., 2011. *Making Enterprise Zones Work: Lessons from Previous Enterpise Zone policy in the United Kingdom,* s.l.: s.n.

UCLA: Statistics Consulting Group, 2020. *Introduction to SAS. UCL: Statistics Consulting Group.* [Online]
Available at: https://stats.idre.ucla.edu/stata/dae/probit-regression/#:~:text=The%20probit%20regression%20coefficients%20give,unit%20change%20in%20the%20predictor.&text=For%20each%20one%20unit%20increase,have%20a%20slightly%20different%20interpretation.
[Accessed 29 June 2020].

United States Census Bureau, 2020. *Cartographic Boundary files - shapefile,* s.l.: s.n.

Van Soest, D., Gerking, S. & Van Oort, F., 2006. Spatial impacts of agglomeration externalities. *Journal of Regional Science,* 46(5), pp. 881-899.

Venables, A. & Storper, M., 2004. Buzz: face-to-face contact and the urban economy. *Journal of Economic Geography,* Volume 4, pp. 351-370.

Verspagen, B. & Schoenmakers, W., 2004. The spatial dimensino of patenting by multinational firms in Europe. *Journal of Economic Geography,* Volume 4, pp. 23-42.

Webb, C., Dernis, H., Harhoff, D. & Hoisl, K., 2005. *Analysing european and international patent citations: a set of EPO patent database building blocks,* s.l.: OECD.

While, A., Jonas, A. & Gibbs, D., 2004. Unblocking the city? Growth pressures, collective provision, and the search for new space of governance in Greater Cambridge, England. *Environment and Planning,* Volume 36, pp. 279-304.

Appendices

Appendix A

Country		Micro region		Macro region		% success
AU	Australia	49	TL3	8	TL2	97
AT	Austria	35	NUTS3	9	NUTS2	99
BE	Belgium	44	NUTS3	3	NUTS2	98
CA	Canada	294	TL3	13	TL2	98
CL	Chile	54	TL3	15	TL2	97
CZ	Czech Republic	14	NUTS3	8	NUTS2	99
DK	Denmark	11	NUTS3	5	NUTS2	99
EE	Estonia	5	NUTS3	1	NUTS2	99
FI	Finland	19	NUTS3	5	NUTS2	99
FR	France	101	NUTS3	27	NUTS2	100
DE	Germany	402	NUTS3	38	NUTS2	97
GR	Greece	52	NUTS3	13	NUTS2	98
HU	Hungary	20	NUTS3	7	NUTS2	96
IS	Iceland	8	NUTS3	2	NUTS2	88
IE	Ireland	8	NUTS3	2	NUTS2	93
IL	Israel	6	TL2	6	TL2	94
IT	Italy	110	NUTS3	21	NUTS2	100
JP	Japan	47	TL3	10	TL2	99
KR	Korea	17	TL3	7	TL2	99
LV	Latvia	6	NUTS3	1	NUTS2	98
LU	Luxembourg	1	NUTS3	1	NUTS2	100
MX	Mexico	209	TL3	32	TL2	94
NL	Netherlands	40	NUTS3	12	NUTS2	99
NZ	New Zealand	14	TL3	14	TL2	97
NO	Norway	19	NUTS3	7	NUTS2	99
PL	Poland	72	NUTS3	16	NUTS2	99
PT	Portugal	25	NUTS3	7	NUTS2	99
SK	Slovak Republic	8	NUTS3	4	NUTS2	99
SI	Slovenia	12	NUTS3	2	NUTS2	97
ES	Spain	59	NUTS3	19	NUTS2	99
SE	Sweden	21	NUTS3	8	NUTS2	100
CH	Switzerland	26	NUTS3	7	NUTS2	99
TR	Turkey	81	NUTS3	26	NUTS2	99
GB	United Kingdom	139	NUTS3	12	NUTS2	98
US	United States	3144	County	51	TL2	98
BR	Brazil			27	TL2	97
BG	Bulgaria	28	NUTS3	6	NUTS2	95
CN	China	35	TL3	34	TL3	97
HR	Croatia	21	NUTS3	2	NUTS2	98
IN	India	36	TL3	36	TL3	97
LT	Lithuania	10	NUTS3	1	NUTS2	87
MT	Malta	2	NUTS3	1	NUTS2	82
RO	Romania	42	NUTS3	8	NUTS2	96
RU	Russian Federation	83	TL3	83	TL3	98
ZA	South Africa	9	TL3	9	TL3	97

(OECD, January 2020)

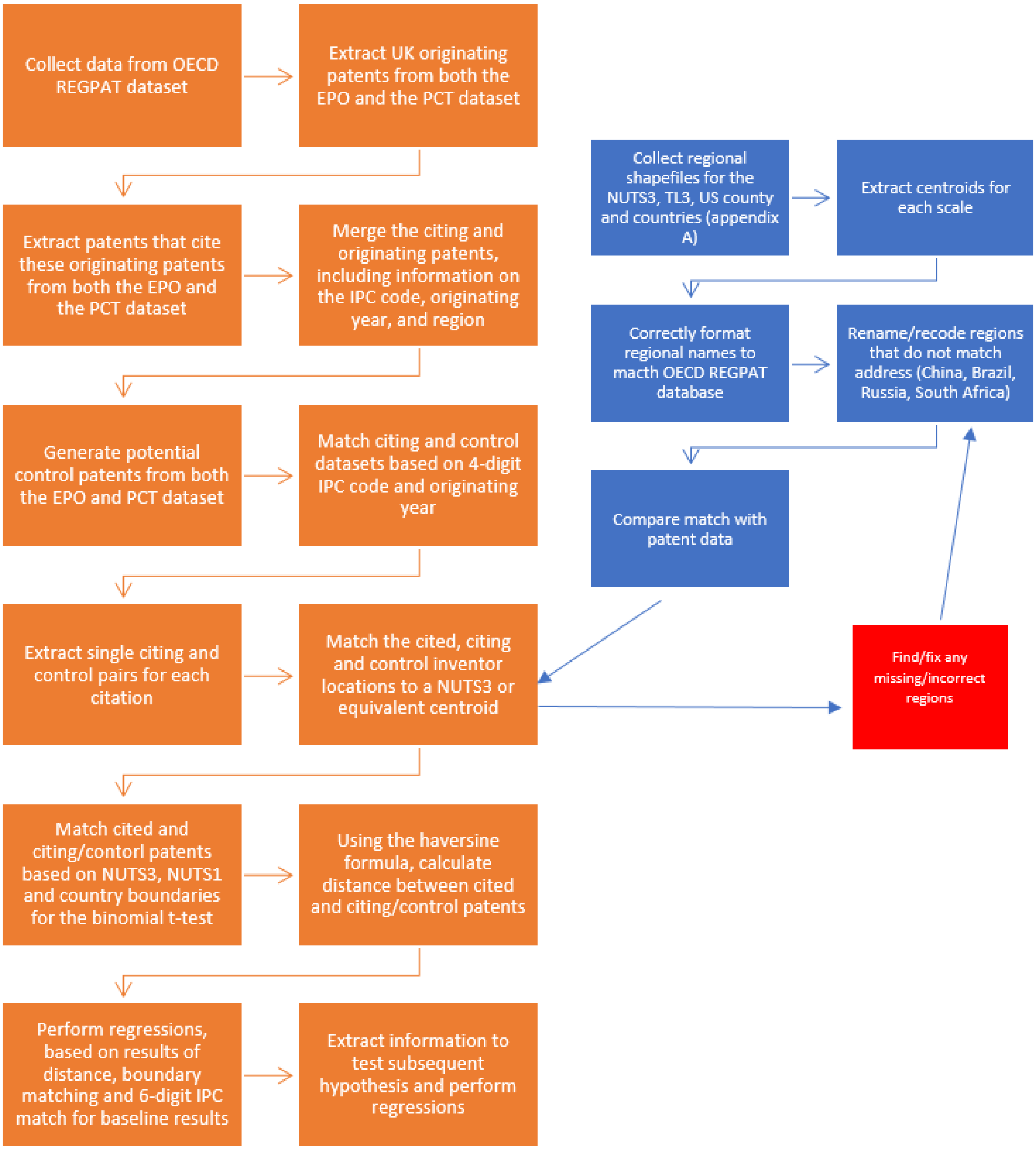
This table was extracted from the OECD REGPAT database accompanying document and details the regional specifications used to assign addresses to micro (local) regions. This was based on the 2013 NUTS3 classification for European countries (excluding England which was based on the 2010 NUTS3 classification), the OECD's Territorial Level 3 units (TL3) and counties for the USA. For countries not covered by these systems, the patent is simply assigned to the country as a single unit (OECD, January 2020).

The data used in this paper for the 2013 and 2010 NUTS3 classifications comes from Eurostat (2020). The data for the US counties data comes from the United States Census Bureau (2020). The data for the OECD TL3 boundaries comes from the OECD (2020). However, non-OECD countries were not covered by the TL3 data, including China, Russia, South Korea, Brazil and South Africa and therefore GADM Unit 1 scales were used (GADM, 2020). To double check how these matched up to the classification used by the OECD REGPAT database, at least 10 addresses for each Non-OECD country (when available) were inputted into google maps and cross referenced compared to the outline of the GADM grouping. The names of the GADM regions were then changed to reflect the distribution of addresses from the OECD database to ensure they matched up to reflect the true distance. This included the performance of over 150 hand renaming of regions for non-OECD regions covered by TL3 specification in the OECD REGPAT database. The results of this are found in the *regional merging.ipynb* file in the GitHub repository. Other country data not covered by any of these specifications was obtained from GADM for which 2 digit country codes also had to be matched to country names and some countries had to be hand identified to reflect the OECD REGPAT database naming convention (GADM, 2020). The workflow for this can be clearly seen in Appendix B.

Appendix B

For the code, please visit: https://github.com/PhilipDW183/Dissertation

For the workflow, please see below.



This workflow represents the workflow of data preparation required in order to be able to run the regressions. This is reflected in the organisation of the GitHub repository which provides detailed commentary on the code and order of workbooks to be able to perform this analysis. Significant amount of time was occupied with aligning regional naming conventions between the OECD REGPAT database and alternate sources of regional information such as GADM as described in Appendix A. This is shown in the red box above in which a circular process was performed by identifying OECD REGPAT regions that had significant numbers of inventors not assigned to a centroid value, entailing the identification of further shapefiles, such as Monaco or the Isle of Man, or the matching of certain micro regions or countries by hand.

Appendix C

The coefficients presented in these table represent the margin effects of a logistic regression model. Thus, they are interpreted as the percentage point change in the probability that the independent variable will equal 1 due a single unit change in the independent variable.

Logistic baseline regression

	Distance (1)	Distance boundaries (2)	Administrative boundaries (3)	Baseline (4)	Top quartile patents (5)	Bottom quartile patents (6)
Boundary match						
Local			0.2936	0.2382	0.1941	0.2547
			(0.000)***	(0.000)***	(0.000)***	(0.000)***
Region			0.1226	0.106	0.1197	0.124
			(0.000)***	(0.000)***	(0.000)***	(0.000)***
Country			0.054	0.0361	0.0464	0.0282
			(0.000)***	(0.000)***	(0.000)***	(0.014)**
Distance (miles)						
ln(distance+1)	-0.0209					
	(0.000)***					
25-50		-0.2785		-0.0992	-0.0317	-0.1297
		(0.000)***		(0.000)***	0.497	(0.000)***
50-100		-0.3496		-0.1194	-0.0121	-0.1268
		(0.000)***		(0.000)***	0.784	(0.000)***
100-250		-0.3812		-0.1047	-0.0141	-0.1174
		(0.000)***		(0.000)***	0.741	(0.000)***
250-500		-0.4295		-0.1275	0.0037	-0.1589
		(0.000)***		(0.000)***	0.931	(0.000)***
500-1000		-0.4329		-0.1289	-0.0046	-0.148
		(0.000)***		(0.000)***	0.914	(0.000)***
1000-2500		-0.4048		-0.1089	0.0683	-0.1847
		(0.000)***		(0.000)***	0.113	(0.000)***
2500-5000		-0.4136		-0.1281	0.0362	-0.1876
		(0.000)***		(0.000)***	0.397	(0.000)***
>5000		-0.4373		-0.1329	0.018	-0.1891
		(0.000)***		(0.000)***	0.673	(0.000)***
Technology match						
IPC 6 match	0.1291	0.1263	0.1247	0.1238	0.0751	0.1492
	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***
R^2	0.01127	0.01482	0.01636	0.01661	0.004703	0.03124
log likelihood	-2.00E+05	-1.99E+05	-1.99E+05	-1.99E+05	-51140	-50140
LLR p-value	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***

Observations	291266	291266	291266	291382	74128	74670

Logistic regression over time

	Original patents prior to 1994 (1)	Original patents after 1994 (2)	Original patents priority year 1977-84 (3)	Original patents priority year 1984-91 (4)	Original patents priority year 1991-98 (5)	Original patents priority year 1998-2005 (6)	Original patents priority year 2005-12 (7)
Boundary match							
Local	0.2360	0.2433	0.1501	0.2620	0.2822	0.2058	0.3534
	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***
Region	0.0924	0.1180	0.0967	0.1026	0.1070	0.1178	0.0849
	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.007)***
Country	0.0420	0.0269	0.0520	0.0382	0.0439	0.0099	0.0181
	(0.002)***	(0.002)***	(0.000)***	(0.001)***	(0.025)**	(0.465)	(0.341)
Distance (miles)							
25-50	-0.0695	-0.1443	-0.1222	-0.0318	-0.1302	-0.1084	-0.1773
	(0.002)***	(0.000)***	(0.001)***	(0.312)	(0.001)***	(0.014)***	(0.003)***
50-100	-0.0698	-0.1916	-0.1014	-0.0602	-0.1389	-0.1669	-0.2295
	(0.001)***	(0.000)***	(0.003)***	(0.049)**	(0.000)***	(0.000)***	(0.000)***
100-250	-0.0740	-0.1500	-0.1086	-0.0499	-0.1164	-0.1524	-0.1612
	(0.000)***	(0.000)***	(0.001)***	(0.088)*	(0.001)***	(0.000)***	(0.002)***
250-500	-0.1041	-0.1661	-0.1383	-0.0762	-0.1362	-0.1687	-0.1867
	(0.000)***	(0.000)***	(0.000)***	(0.010)**	(0.000)***	(0.000)***	(0.000)***
500-1000	-0.1138	-0.1620	-0.1537	-0.0831	-0.1411	-0.1659	-0.1680
	(0.000)***	(0.000)***	(0.000)***	(0.005)***	(0.000)***	(0.000)***	(0.002)***
1000-2500	-0.0965	-0.1423	-0.1594	-0.0611	-0.1241	-0.1280	-0.1777
	(0.000)***	(0.000)***	(0.000)***	(0.044)**	(0.001)***	(0.000)***	(0.001)***
2500-5000	-0.1186	-0.1576	-0.1696	-0.0894	-0.1304	-0.1646	-0.1656
	(0.000)***	(0.000)***	(0.000)***	(0.003)***	(0.000)***	(0.000)***	(0.002)***
5000+	-0.1232	-0.1665	-0.1589	-0.0955	-0.1451	-0.1720	-0.1712
	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***
Technology match							
IPC 6 match	0.1470	0.1059	0.1709	0.1380	0.1059	0.1090	0.0966
	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***
R^2	0.0247	0.0115	0.03414	0.02353	0.01267	0.01074	0.01231
log likelihood	-90442	-1.15E+05	-27978	-48157	-50925	-43151	-24993
LLR p-value	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***
Observations	133784	153004	41790	71150	74412	62930	36506

Logistic regression for knowledge dissemination

	Citation lag less than 6 years (1)	Citation lag longer than 6 years (2)	Bottom 25% for citation lag (3)	25-50% for citation lag (4)	50-75% for citation lag (5)	Top 25% for citation lag (6)
Boundary match						
Local	0.2621	0.1976	0.2727	0.2485	0.1713	0.2279
	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***
Region	0.0974	0.1189	0.1042	0.0876	0.1220	0.1145
	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***
Country	0.0459	0.0245	0.0603	0.0271	0.0087	0.0430
	(0.000)***	(0.005)***	(0.000)***	(0.024)**	(0.459)	(0.001)***
Distance (miles)						
25-50	-0.1092	-0.0810	-0.1072	-0.1106	-0.1171	-0.0324
	(0.000)***	(0.004)***	(0.000)***	(0.002)***	(0.002)***	0.458
50-100	-0.1267	-0.1045	-0.1300	-0.1197	-0.1565	-0.0339
	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	0.405
100-250	-0.1154	-0.0848	-0.1176	-0.1108	-0.1194	-0.0398
	(0.000)***	(0.001)***	(0.000)***	(0.001)***	(0.000)***	0.307
250-500	-0.1389	-0.1080	-0.1293	-0.1497	-0.1646	-0.0363
	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	0.354
500-1000	-0.1412	-0.1095	-0.1318	-0.1524	-0.1687	-0.0354
	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	0.365
1000-2500	-0.1123	-0.0990	-0.1038	-0.1219	-0.1511	-0.0327
	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	0.407
2500-5000	-0.1237	-0.1268	-0.1023	-0.1505	-0.1793	-0.0598
	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	0.126
5000+	-0.1338	-0.1251	-0.1134	-0.1589	-0.1743	-0.0614
	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***
Technology match						
IPC 6 match	0.1151	0.1323	0.1077	0.1237	0.1346	0.1296
	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***
R^2	0.01871	0.01471	0.01943	0.0182	0.01615	0.01346
log likelihood	-1.05E+05	-93971	-59745	-44757	-48096	-45864
LLR p-value	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***
Observations	153670	137596	87902	64768	70526	67070

Logistic regression for different industries

	Biotechnology	Medical	Pharmaceuticals	Chemicals (exc. Pharmaceuticals)	Engineering	IT	Telecommunications	Electronics
Boundary match								
Local	0.3427	0.3306	0.3584	0.1869	0.234	0.2631	-0.0452	0.2515
	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	0.690	(0.000)***
Region	0.1620	0.0973	0.0700	0.1361	0.1127	0.0528	0.0758	0.0389
	(0.002)***	(0.020)**	(0.050)*	(0.000)***	(0.000)***	(0.064)*	0.211	0.186
Country	0.0603	0.0345	0.0276	0.0417	0.0672	0.0294	0.0914	0.0253
	(0.031)**	0.162	0.172	(0.000)***	(0.000)***	(0.062)*	(0.006)***	0.123
Distance (miles)								
25-50	-0.0242	-0.0456	-0.0700	-0.1049	-0.1183	-0.0621	-0.2044	-0.1591
	0.831	(0.056)*	0.276	(0.002)***	(0.005)***	0.223	0.103	(0.003)***
50-100	-0.1089	-0.0167	-0.0962	-0.1007	-0.1221	-0.1109	-0.3043	-0.1520
	0.308	0.825	0.128	(0.002)***	(0.002)***	(0.024)**	(0.012)**	(0.004)***
100-250	-0.0522	-0.0310	-0.1171	-0.0808	-0.1266	-0.0917	-0.1940	-0.2210
	0.616	0.669	(0.052)*	(0.009)***	(0.001)***	(0.051)*	(0.094)*	(0.000)***
250-500	-0.0293	-0.0308	-0.1782	-0.1073	-0.1311	-0.1256	-0.2262	-0.2435
	0.778	0.677	(0.003)***	(0.001)***	(0.001)***	(0.008)***	(0.052)*	(0.000)***
500-1000	-0.0303	-0.0390	-0.1459	-0.1025	-0.1190	-0.1219	-0.2142	-0.2233
	0.772	0.598	(0.016)**	(0.001)***	(0.002)***	(0.010)**	(0.066)*	(0.000)***
1000-2500	0.0157	-0.0246	-0.1092	-0.0911	-0.1108	-0.1037	-0.2050	-0.2370
	0.881	0.742	(0.074)*	(0.004)***	(0.004)***	(0.030)**	(0.081)*	(0.000)***
2500-5000	0.0157	-0.0391	-0.1161	-0.1076	-0.1169	-0.1298	-0.2377	-0.2521
	0.880	0.596	(0.054)*	(0.001)***	(0.002)***	(0.006)***	(0.040)**	(0.000)***
5000+	-0.0169	-0.0358	-0.1087	-0.1144	-0.1332	-0.1172	-0.2385	-0.2488
	0.871	0.628	(0.071)*	(0.000)***	(0.000)***	(0.013)**	(0.040)**	(0.000)***
Technology match								
IPC 6 match	0.0629	0.1577	0.0838	0.1296	0.1178	0.0855	0.0669	0.1014
	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.009)***
R^2	0.01014	0.01766	0.01936	0.01576	0.02707	0.009344	0.006318	0.01255
log likelihood	-7728.5	-12292	-12708	-51871	-35381	-28806	-6649.4	-27684
LLR p-value	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***
Observations	11264	18052	18696	76032	52464	41950	9654	40435

Appendix D

(OECD.stat, 2020)

Biotech: ["A01H001", "A01H004", "A01K067", "A01K035/12", "A01K035/13", "A01K035/14","A01K035/15", "A01K035/16", "A01K035/17", "A01K035/18", "A01K035/19", "A01K035/20", "A01K035/21", "A01K035/22", "A01K035/23", "A01K035/24", "A01K035/25", "A01K035/26", "A01K035/27", "A01K035/28", "A01K035/29", "A01K035/30", "A01K035/31", "A01K305/32", "A01K035/33", "A01K035/34", "A01K035/35", "A01K035/36", "A01K035/37", "A01K035/38", "A01K035/39", "A01K035/40", "A01K035/41", "A01K035/42", "A01K035/43", "A01K035/44", "A01K035/45", "A01K035/46", "A01K035/47", "A01K035/48", "A01K035/49", "A01K035/50", "A01K035/51", "A01K035/52", "A01K035/53", "A01K035/54", "A01K035/55", "A01K035/56", "A01K035/57", "A01K035/58", "A01K035/59", "A01K035/60", "A01K035/61", "A01K035/62", "A01K035/63", "A01K035/64", "A01K035/65", "A01K035/66", "A01K035/67", "A01K035/68", "A01K035/69", "A01K035/70", "A01K035/71", "A01K035/72", "A01K035/73", "A01K035/74", "A01K035/75", "A01K035/76", "A01K035/77", "A01K035/78", "A01K035/79", "A61K38", "A61K039", "A16K048", "C02F003/34", "C07G011", "C07G013", "C07G015", "C07K004", "C07K014", "C07K016", "C07K017", "C07K019", "C12M", "C12N", "C12P", "C12Q", "C40B010", "C40B040/02", "C40B040/03","C40B040/04", "C40B040/05", "C40B040/06", "C40B040/07", "C40B040/08", "C40B050/06", "G01N027/327", "G01N033/53", "G01N033/54", "G01N033/55", "G01N033/57", "G01N033/68", "G01N033/74", "G01N033/76", "G01N033/78", "G01N033/88", "G01N033/92", "G06F019/10","G06F019/11", "G06F019/12", "G06F019/13", "G06F019/14", "G06F019/15", "G06F019/16", "G06F019/17", "G06F019/18", "G06F019/20", "G06F019/21", "G06F019/22", "G06F019/23", "G06F019/24"]

(OECD.stat, 2020)

Pharmaceutical: ["A61K"]

(OECD.stat, 2020)

Medical: ["A61B", "A61C", "A61D","A61F", "A61G", "A61H", "A61J", "A61L", "A61M", "A61N","H05G"]

(Inaba & Squicciarini, 2017)

IT: ["H03K", "H03L", "H03M", "H04B001", "H04J", "H04L", "H04M003", "H04M004", "H04M005", "H04M006", "H04M007", "H04M008", "H04M009", "H04M010", "H04M011", "H04M012", "H04M013", "H04M019", "H04M099", "H04Q", "H04B001/00","H04B001/01", "H04B001/02", "H04B001/03", "H04B001/04", "H04B001/05", "H04B001/06","H04B001/07", "H04B001/08", "H04B001/09", "H04B001/10", "H04B001/11", "H04B001/12", "H04B001/13", "H04B001/14", "H04B001/15", "H04B001/16", "H04B001/17", "H04B001/18", "H04B001/19", "H04B001/20", "H04B001/21", "H04B001/22", "H04B001/23", "H04B001/24", "H04B001/25", "H04B001/26", "H04B001/27","H04B001/28", "H04B001/29", "H04B001/30", "H04B001/31", "H04B001/32", "H04B001/33", "H04B001/34","H04B001/35", "H04B001/36", "H04B001/37", "H04B001/38", "H04B001/39", "H04B001/40","H04B001/41","H04B001/42", "H04B001/43", "H04B001/44", "H04B001/45", "H04B001/46", "H04B001/47", "H04B001/48","H04B001/49", "H04B001/50", "H04B001/51", "H04B001/52", "H04B001/53", "H04B001/54", "H04B001/55", "H04B001/56", "H04B001/57", "H04B001/58", "H04B001/59", "H04B001/60", "H04B001/61", "H04B001/62","H04B001/63", "H04B001/64", "H04B001/65", "H04B001/66", "H04B001/67", "H04B001/68", "H04B001/69", "H04B001/72", "H04B001/73", "H04B001/74", "H04B001/75", "H04B001/76", "H04B003", "H04B004", "H04B008", "H04B006", "H04B009", "H04B010", "H04B011",

"H04B012", "H04B013", "H04B014", "H04B015","H04B016", "H04B017","H04H", "H04B007", "H04W", "G06F012/14", "G06F021", "G06K019", "G09C", "G11C008/20", "H04K","H04L009", "H04M001/66", "H04M001/67", "H04M001/68", "HM04M001/70", "H04M1/727","H04N007/167", "H04N007/169", "H04N007/171", "H04W12", "C06Q020", "G07F007/08", "G07F007/09", "G07F007/10", "G07F007/11", "G07F007/12", "G07G001/12", "G07G001/13","G07G001/14", "H04L012/14", "H04W004/24", "G08B001/08","G08B003/10", "G08B005/22", "G08B005/23", "G08B005/24", "G08B005/25","G08B005/26","G08B005/27", "G08B005/28", "G08B005/29", "G08B005/30","G08B005/31","G08B005/32","G08B005/33","G08B005/34", "G08B005/35", "G08B005/36", "G08B005/37","G08B005/38", "G08B007/06", "G08B013/18", "G08N013/19", "G08B025", "G08B026", "G08B027", "G08C", "G08G001/01", "G08G001/02", "G08G001/04", "G08G001/05", "G08G001/065", "G06F017/40", "H04W084/18", "H04B001/59", "H04B005", "G01S013/74", "G01S013/75", "G01S013/76", "G01S013/77", "G01S013/78", "G01S013/79", "G01S013/80", "G01S013/81","G01S013/82","G01S013/84", "G01V003", "G01V015", "H04W084/10", "G06F005", "G06F007", "G06F009", "G06F011","G06F013", "G06F015/00", "G06F015/16","G06F015/17","G06F003/06","G06F003/07","G06F003/08", "G06F012","G06K001","G06K002","G06K003","G06K004","G06K005","G06K006","G06K007","G06K013", "G11B", "G11C","H04N005/78", "H04N005/80" ,"H04N005/82", "H04N005/82", "H04N005/83" ,"H04N005/84", "H04N005/85","H04N005/87","H04N005/89","H04N005/90", "G06F017/30", "G06F017/40", "G06F017/00", "G06F017/10", "G06F017/11", "G06F017/12", "G06F017/13", "G06F017/14", "G06F017/15","G06F017/16","G06F017/17","G06F017/18", "G06F019", "G06Q010", "G06Q030", "G06Q040", "G06Q050", "G06Q090", "G06Q099", "G08G", "G06F017/20", "G06F017/21","G06F017/22","G06F017/23","G06F017/24","G06F017/25","G06F017/26","G06F017/27", "G06F017/28", "G06K009", "G06T007", "G10L013/027", "G10L015", "G10L017", "G10L025/63", "G10L025/66", "G06F015/18", "H04M001", "G06F003/01", "G06F003/02", "G06F003/03", "G06F003/04", "G06F003/14","G06F003/15","G06F003/16","G06K011", "G08G001/0962", "G08G001/0963","G08G001/0964","G08G001/0965","G08G001/0966","G08G001/0967","G08G001/0968","G08G001/0969","G09B005", "G09B007", "G09B009", "G06F017/50", "G06K009", "G06T011", "G06T013", "G06T015""G06T017", "G06T018", "G06T019","H04N", "G06T001", "G06T002", "G06T003", "G06T004","G06T005""G06T006","G06T008","G06T009", "G09G", "H04R", "H04S", "G10L", "H03B", "H03D", "H03F", "H03G", "H03H", "H03J", "H01B011", "H01L029","H01L030", "H01L031", "H01L032", "H01L033", "H01L021","H01L025","H01L027","H01L043","H01L044" "H01L045","H01L046","H01L047","H01L048","H01L049","H01L050","H01L051", "G02B6", "G02F", "H01S0005", "B81B007/02", "B82Y010", "H01P", "H01Q", "G01S", "G01V003", "G01V008", "G01V015", "G06F003/00", "G06F003/05", "G06F003/09", "G06F003/12", "G06F003/13", "G06F003/18", "G06E", "G06F001", "G06F0015/02", "G06F015/04", "G06F015/08", "G06F015/09","G06F015/10","G06F015/11","G06F015/12","G06F015/13", "G06F015/14", "G06G007", "G06J", "G06K015", "G06K017", "G06N", "H04M015", "H04M017"]

Excluding: ["H04L009", "H04L012/14", "H04B001/59", "H04B005", "H04B007", "H04W004/24", "H04W012","G06T011/80", "GC11C008/20", "G08G001/01", "G08G001/02", "G08G001/03", "G08G001/04", "G08G001/05", "G08G001/06", "G08G001/07", "G08G001/08", "G08G001/09", "G08G001/010", "G08G001/011","G08G001/012", "G08G001/013", "G08G001/014", "G08G001/015", "G08G001/016", "G08G001/017","G08G001/018","G08G001/019", "G08G001/020", "G08G001/021", "G08G001/022", "G08G001/023", "G08G001/024", "G08G001/025", "G08G001/026", "G08G001/027", "G08G001/028", "G08G001/029", "G08G001/030", "G08G001/031","G08G001/032","G08G001/033", "G08G001/034", "G08G001/035", "G08G001/036","G08G001/037","G08G001/038","G08G001/039","G08G001/040" "G08G001/041","G08G001/042","G08G001/043","G08G001/044","G08G001/045","G08G001/046"," G08G001/047","G08G001/048", "G08G001/049", "G08G001/050", "G08G001/051", "G08G001/052","G08G001/053","G08G001/054", "G08G001/055", "G08G001/056", "G08G001/057", "G08G001/058", "G08G001/059", "G08G001/060","G08G001/061", "G08G001/062", "G08G001/063", "G08G001/064", "G08G001/065", "G08G001/0962", "G08G001/0963", "G08G001/0964", "G08G001/0965", "G08G001/0967", "G08G001/0968", "G08G001/0969"]

Belenzon and Schankerman (2013)

Chemicals: ["C0", "C1", "B01", "D01F", "A62D"]

Engineering: ["A01B", "A01C", "B021", "D21", "B06B", "B09", "B21", "B22", "B23", "B25", "B29", "B60", "B62", "B65", "B81", "B82", "D01D", "D02", "D03", "D04","D05", "D06M", "D21", "E21", "F04", "F25", "G05G", "G07"]

Electronics: ["H01L", "H03", "G11C", "G06C", "G06D", "G06E", "G06F011", "G06F015", "G06F017", "G06G", "H01", "H02", "H04", "H05", "B03C"]

Telecommunications: ["H04L", "H04M", "H04N"]

Research log

Date	Task	Challenges	Solutions
1/05	Initial exploration of the literature in the first week of term	Finding a large corpus of literature in the area that is not necessarily relevant	Narrow down exploration to only those highly relevant to topic i.e. knowledge spillovers and use of patents
6/05	Finding data to match with questions that can be asked following initial exploration of the literature	UK patent applications cannot be easily accessed from the UKIPO	Use OECD REGPAT data which combines EPO and PCT datasets that can be linked to the UK and hence specifically follow UK knowledge spillovers. This may allow for further exploration later.
15/05	Initial draft of literature review to be produced	Fitting everything into a coherent story that will inform the research questions and that is not too long	Put all the work into a chronological order and write as much as possible so that it can later be adapted to the specific research question explored through the data obtained
22/05	First meeting with supervisor over work produced	Producing a literature review that covers all the areas that are expected to be covered based on reading so far.	All the key works referenced, and a significant corpus of literature covered, it is now about using the data to answer specific research questions relevant to the body of literature explored
22/05	Initial exploration of the dataset	Matching citations across both the EPO and the PCT dataset, extracting relevant information for later regressions and finding a control sample	Work over this week primarily consists of producing the cited, citing and control samples. The code is produced to extract all of these from the

			existing data in python and R.
29/05	Matching patents to a geographic location	Post codes could not be extracted, most of the citations/control group were not from the UK, the use of a dataset that has exact locations results in a large loss of data	Use the NUTS3 localities assigned by the OECD REGPAT database and find shapefiles that correspond to those used by the OECD REGPAT database
5/06	Matching patents to a geographic location (part 2)	Some of the shapefiles don't cover the necessary areas or the regions do not match up	Spend some time looking at addresses and linking these to regions with different regional codes. Find additional shapefiles that are required. Explore nuances within the dataset that do not match to find solutions
12/06	Performing regression results	Previous papers have used a variety of regression methodologies	Use Probit regression because of its prevalence within the econometric literature and the results are the same as logistic regression ones. Multiple regressions are performed for each specification and the results are compared, along with their strengths and weaknesses
19/06	Write up results into a rough draft	There are a lot of different areas to talk about and a lot that can be said about each result	Use this as a rough draft to get all the important points down first. This will be significantly over the word count, but it can be cut down/cleaned up as the final draft develops

23/06	Meeting with supervisor to discuss results	Methodology is good but need more clarity, and results need figures to provide graphical explanation of tables. Some of the wording in the results is redundant and so needs to be removed. Figure captions need to be added.	Methodology rewritten and cleaned up, with line graphs produced of results. Results are shortened and cleaned up so that they follow a clearer narrative and are directly related to the hypotheses created. Figure captions are added so that they are made clearer.
03/07	First complete draft produced	Significantly over the word count and the order is not logical in some areas	Begin to narrow down the focus and rewrite the literature review to align with the data and questions asked/answered. Conclusion and introduction are rewritten and made more succinct and relevant.
17/07	Meeting with supervisor to discuss cleaned up draft	Issues relating to methodology/data order and clarity, along with lack of explanation of limitations/hypothesis and word count	Literature review rewritten to explicitly incorporate hypotheses to be explored, data and methodology re-ordered and merged to create two clearly separate sections, excess areas trimmed to reduce the word count
27/07	Updated draft sent to supervisor		Word count has been reduced, appendices have been added, GitHub repo linked to, prior issues resolved
09/08	Final piece of work submitted		All comments resolved as much as possible,

	presentation is cleaned up, grammatical errors checked.