Journal of Urban Economics 112 (2019) 16–32

Contents lists available at ScienceDirect

Journal of Urban Economics

journal homepage: www.elsevier.com/locate/jue

London calling? Agglomeration economies in literature since 1700

Sara Mitchell

TU Dortmund, Germany

ARTICLE INFO

JEL classification:
D24
J24
J61
N90
R11
Z11

Keywords:
Geographic concentration
Cities
Mobility
Productivity
Urban history
British literature

ABSTRACT

This paper utilises a unique, purpose-built panel dataset on prominent authors in the UK and Ireland born 1700–1925 to estimate the productivity gains associated with agglomeration of an industry with few capital requirements and no apparent need to cluster geographically. I find the average author experiences productivity gains of 11.94% per annum when residing in London, the only major literary cluster – a gain not associated with living in any of the minor literary clusters. I find evidence of negative selection with respect to productivity, indicating the results are not driven by the self-selection of highly productive authors to London. I find heterogeneity of returns to living in London by birth cohort and Impact Index quartile (a measure of author quality) and that the cohorts who receive the greatest gains from locating in London are those for which there is the strongest evidence of negative selection with respect to productivity.

1. Introduction

The potential for gains from agglomeration of industry is evident for industries with traditional modes of production. However, less is known empirically about these agglomeration effects for industries with few capital requirements and no apparent need to cluster geographically, such as literary production. Unlike a factory or science laboratory, writing does not require a centralised location of production. There are few capital requirements for the writing of literature, and the printing process does not require the author to be present. Authors are able to send drafts and conduct most business with a publisher or literary agent through the post or, in more recent years, via email. The data also suggests authors are highly mobile. Given these unique characteristics, do authors tend to cluster geographically at all? If so, are there gains in productivity associated with the agglomeration of literary activity?

In general, the geographic concentration of industry occurs because firms benefit from economies of scale via increased input-output linkages, labour market pooling, and technology spillovers (Krugman, 1991; Duranton and Puga, 2004; Combes and Gobillon, 2015). As Alfred Marshall explained:

> *"When an industry has thus chosen a locality for itself, it is likely to stay there long: so great are the advantages which people following the same skilled trade get from near neighbourhood to one another. The mysteries of the trade become no mysteries; but are as it were in the air, and children learn many of them unconsciously. Good work is rightly appreciated... if one man starts a new idea, it is taken up by others and combined with suggestions of their own; and thus it becomes the source of further new ideas... "* (Marshall, 1890, p. 225)

In recent years, gains from agglomeration have been observed across a range of creative and innovative activities. Perhaps the most prevalent method of studying the gains from geographic proximity of creative and innovative activity is by "following the paper trail" in analysing the spatial clustering of patent registrations from R&D by private firms and universities.[1] As Audretsch and Feldman (2004) note, "the basic results agree: Patent citations are highly localized, indicating that location and proximity clearly matter in exploiting knowledge spillovers."

Another set of literature focuses on academic research. Waldinger (2012a) finds the breakdown of research clusters (through the dismissal of Jewish scientists in Nazi Germany) results in significant decreases in departmental productivity. Waldinger (2012b, 2010) find that higher cluster density and higher quality of peers within the cluster both increase the productivity of researchers and have a positive effect on the outcomes of the Ph.D. students within their department.

Another body of literature looks at the agglomeration of cultural production. Borowiecki (2013) finds that the clustering of classical music

[1] See Carlino et al., 2007; Jaffe et al., 1993; Moser et al., 2014; Audretsch and Feldman, 2004; Azoulay et al., 2010; Audretsch and Feldman, 1996, among others.

E-mail address: sara.mitchell@tu.dortmund.de

https://doi.org/10.1016/j.jue.2019.05.002
Received 12 January 2016; Received in revised form 3 May 2019
Available online 15 May 2019

composers has a positive and significant impact on composer productivity. Borowiecki (2012) and Borowiecki and O'Hagan (2013) find the outward migration of composers during periods of war significantly reduces a country's long-term creative potential. Hellmanzik (2010) finds that the artwork of prominent visual artists is valued higher when produced in a creative cluster. Giorcelli and Moser (2015) find that copyrights lead to an increase in both the number of Italian operas and the number of high-quality Italian operas, and they find that cities with pre-existing industry-related infrastructure received greater gains.

However, these industries have capital requirements that motivate agglomeration. The critical question in this context is thus: is literary production different? With the ability to submit writings by post, where would authors choose to live and work? Is the average author like the Brontë sisters who are often styled as solitary geniuses – or more like the Lost Generation of Paris in the 1920s, a dense network of artists, musicians, writers, and friends? How might the Marshallian agglomeration mechanisms operate in the context of literary production, if at all?

In order to answer these questions, I determine the cities in which literary activity clusters, and I empirically estimate the returns to literary agglomeration. Specifically, I utilise a unique, purpose-built dataset with information on the birth location and lifetime migration, productivity in terms of number of publications, and demographic characteristics of 370 authors in the UK and Ireland since 1700. I begin by discussing the data collection methodology. I then analyse the patterns of migration and clustering (i.e. geographic co-location) of authors, and I construct age-productivity profiles to determine the productivity gains (if any) associated with living in a literary cluster.

I find that London emerges as the only major literary cluster, with over 50% of authors locating in London in a given year during the periods with the greatest concentration of authors. In comparison, less than 20% of the UK population resided in Greater London in any given Census year. I then estimate the "London effect" utilising individual fixed effects to control for spatial sorting by skill level, and the within estimator reveals the average author annual productivity is 11.94% per annum higher when the author resides in London. Similar agglomeration gains are not observed for the minor literary clusters. I find evidence of negative selection with respect to productivity, indicating the results are not driven by the self-selection of highly productive authors to London. I find heterogeneity of returns to living in London by birth cohort and Impact Index quartile (a measure of author quality), and I find that the cohorts who receive the greatest gains from locating in London are those for which there is the strongest evidence of negative selection with respect to productivity (authors born in the 18th century and authors ranking in the top two quartiles of the Impact Index).

The remainder of this paper is structured as follows. In Section 2, I discuss the mechanisms responsible for innovative activity (using the Duranton and Puga, 2004, taxonomy) and how these might be realised in the context of literary production. Next, I discuss the data collection methodology and descriptive statistics in Section 3. Section 4 outlines the identification strategy, and results are presented in Section 5. Section 6 concludes.

2. Agglomeration economies of literary production

Innovative activity can be facilitated by thick local markets that enable more efficient sharing of inputs and provide a larger pool of specialised workers and services (Carlino and Kerr, 2015), a mechanism referred to as *sharing* in the Duranton and Puga (2004) taxonomy. While increased concentration of industry can increase the number of new ideas that are generated, "it is possible that concentration encourages innovation by assisting in the realization of ideas rather than in (or perhaps in addition to) the generation of ideas" (Helsley and Strange, 2002). In markets with dense networks of input suppliers, the process of input sharing reduces the cost of realising ideas.

In the context of literary production, the sharing of inputs and other resources within the book publishing and manufacturing industry could lead to efficiencies in the production process, enabling more unique works to be published and more copies of each work to be printed. Thus, authors would choose to live in the cities that enable the greatest gains in productivity (where they have the highest probability of getting a work published) – i.e. the cities with the greatest concentration of book publishing and manufacturing firms.

In the British Isles, the geographic concentration of the publishing industry was influenced by regional differences in copyright enforcement. The first copyright law was introduced in 1710 when the British parliament passed the *Statute of Saint Anne*.[2] While this Statue legally applied in both Scotland and England, the English publishers acted as a cartel, maintaining their monopoly over English publishing and continuing to uphold the previous system of perpetual copyright. Scottish courts, however, actively enforced the new copyright law.

This divergence in the enforcement of copyright allowed for a dramatic expansion of the publishing and book manufacturing industry in Scotland, particularly in Edinburgh.[3,4] Not only were the Scottish printers able to produce books at lower cost, they were also able to sell books in England – even as far south as Oxford and Cambridge – on credit for prices "three or more times more favourable than those offered by London" (St Clair, 2004, p. 106–107). By the 1750s, the Scottish printing industry also offered lower prices on printing equipment.

Although London remained the largest market, Edinburgh was a competitive location. Even as the London cartel's monopoly weakened, both Edinburgh and London remained the main book manufacturing centres for the Romantic and Victorian periods and into the 20th century (St Clair, 2004). Therefore, if the sharing mechanism is important for literary production, then one would expect to see authors distributed between London, Edinburgh, and the rest of the UK and Ireland in a way that reflects the relative availability of the local inputs of the book publishing and manufacturing industries.

[2] The *Statute of Saint Anne*, also known as the *Copyright Act of 1709*, mandated that all books be given a copyright of 14 years if published after 10 April 1710 or 21 years if published before. Before the Statute of Saint Anne, a commercial book market primarily based in London flourished under a system of perpetual copyright. This system of perpetual copyright was regulated and enforced by the Stationers' Company, a London-based guild that held monopoly power over the publishing of books in the United Kingdom. See St Clair (2004) and the *'History of Publishing'* and '*Copyright*' sections in Encyclopaedia Britannica (2014) for more information.

[3] St Clair (2004) notes: "At the time Scottish judges made their first decisions, there was no substantial book industry in Scotland, nor any powerful constituency of intellectual property owners. The effect was therefore not only to create a difference in what was legally permitted in the two countries of the union, but to set up economic incentives for the development of a Scottish book industry, as a kind of offshore rival to England but without any of the usual legal, physical or tariff barriers... To reprint a title which had passed into the public domain under the terms of the 1710 Act was not illegal, nor subject to injunction, in Scotland, nor was it illegal to sell such books in England or in the British colonies in America and elsewhere... By the middle of the eighteenth century, a substantial Scottish book industry had grown up in Edinburgh, and others in Glasgow and Aberdeen, publishing a variety of newly written texts under statutory copyright, but mainly reprinting works which lay outside the limits of the 1710 Act. The Scottish publishers were able to do this profitably at prices which were between half and two-thirds of the levels of the London industry." (p. 105)

[4] From the date copyright was introduced, the London booksellers and printers fought the copyright law in court and appealed to parliament for a repeal, eventually engaging in a 30-year legal 'battle of the booksellers' that ended in 1774 with the *Donaldson v. Becket* judgment upholding statutory copyright limits (Rose, 1988). However, St Clair (2004) notes that "the London publishers continued to deal in legally unenforceable perpetual intellectual properties for at least forty years after 1774" (p. 111). While there were other changes in copyright law through the 19th and 20th centuries, none resulted in further major divergences between Scottish and English publishing industries, as the London cartel's monopoly power over copyright continued to weaken.

Having access to a larger pool of resources not only enables the sharing of inputs but also improves the quality of matches, a mechanism known as *matching* in the Duranton and Puga (2004) taxonomy. Carlino and Kerr (2015) explain that the availability of opportunities in thick local markets reduces the opportunity cost of waiting for a partner, so workers can be more selective in forming matches. Workers are also more mobile, since the concentration of innovative activity makes it easier for workers to find new positions. This results in higher quality matches, increasing the productivity of matches and enabling a greater proportion of the labour force to be involved in productive matches.

In the context of literary production, the agglomeration of literature and related cultural industries would increase the probability of attaining high-quality matches across social and professional networks, including publishers, editors, literary critics, intellectuals and others who were prominent or influential in British and Irish society. For example, having access to a large pool of publishing houses would enable authors to be selective in choosing which publishing house to work with and allow authors to avail of other opportunities if they are unhappy with a contract with a particular publisher. Having access to high-quality networks could have other benefits, such as enabling introductions to industry 'gate-keepers', ensuring that a new work was shared with publishers or literary critics, and allowing access to information about the latest literary or cultural trend.

A famous example of such a network in British and Irish literary history is the Bloomsbury group, a diverse group of influential writers, artists, and intellectuals, many of whom lived near one another in the Bloomsbury district in London (Encyclopaedia Britannica, 2014). This informal group included Virginia Woolf (writer), Leonard Woolf (writer), E.M. Forster (writer), Clive Bell (art critic), Vanessa Bell (painter), Roger Fry (art critic and painter), Duncan Grant (painter), Desmond MacCarthy (literary journalist), Lytton Strachey (biographer), and John Maynard Keynes (economist).[5] Their wider social circle included a number of other influential individuals, including authors such as T.S. Eliot, Katherine Mansfield, Hugh Walpole, and Vita Sackville-West.

Not only does the geographic concentration of industry increase the probability of having higher quality matches, it also enables information and knowledge to flow through these connections – a mechanism known as *learning* in the Duranton and Puga (2004) taxonomy. More specifically, Duranton and Puga highlight several types of learning mechanisms that are particularly relevant in the context of creative production: knowledge generation, diffusion and accumulation. Their discussion of these mechanisms alludes to two important concepts – that of 'space of places' (the importance of location for learning and innovation) and that of 'space of flows' (the importance of networks in the transfer and diffusion of knowledge) (Ter Wal and Boschma, 2009). The importance of these concepts in learning-based agglomeration economies becomes clear with a careful understanding of the difference between information and knowledge, particularly with regard to tacit knowledge.

Tacit knowledge is gained through an individual's perception of information through repeated interaction in a shared learning process; this contrasts with explicit knowledge, which is information that can be codified and stored in media such as an encyclopaedia or a textbook (Howells, 2012). Tacit knowledge is transmitted most efficiently between individuals in close proximity, so there are natural geographic boundaries to its flows and spillovers.[6] Thus, it is natural that "innovative activity should concentrate geographically in those industries where the direct knowledge-generating inputs are greatest and where knowledge spillovers are the most prevalent" (Audretsch and Feldman, 1996).[7,8]

Authors are often described as solitary and independent workers.[9] While it is possible that authorship is an inherently individual process and thus authors would not benefit to the same degree (if at all) from agglomeration, there are also examples of close collaboration and learning within literary circles. For example, Joseph Conrad and Ford Madox Ford worked very closely for several years, reading and editing one another's works, exchanging ideas, and even co-authoring several works. Farrell (2003) explains that "Conrad knew the basic 'architectonics' of storytelling, while Ford contributed his stylistic skills and his ear for English " (p. 134) and that "... because of their extensive collaboration, Ford was able to take up where Conrad left off and almost seamlessly carry the story and characters forward" (p.134). This experience had clear benefits for both authors: "several critics and biographers have argued that most of Conrad's best work was done while he was paired in this collaborative friendship with Ford" (Farrell, 2003, p. 132).

Therefore, despite some anecdotal evidence suggesting the contrary, all of the Duranton and Puga (2004) agglomeration mechanisms have the potential to be realised within literary agglomerations. There are also a number of other agglomeration forces that may be at work in this context. For example, London may have offered better employment opportunities for authors who were not successful enough to earn a sufficient income through writing alone. London may also have offered a diversity of experiences that were useful to the writing process. It is also possible that the many London-based readers enjoyed reading stories about their own city, requiring authors to spend at least some time in the city to give accurate portrayals. Authors may also have simply enjoyed proximity to the reading public of London.[10]

Regardless of the underlying agglomeration forces, it is clear that, even for an industry with few capital requirements or with no apparent

[5] John Maynard Keynes' involvement with the Bloomsbury group had a lasting impact not only on him personally but also on British art and society: Keynes helped to found the Council for the Encouragement of Music and the Arts in 1946 (later renamed the Arts Council of Great Britain) and became the first Chair of the Council (Upchurch, 2004).

[6] See Audretsch (1998); Gertler (2003); Von Hippel (1994).

[7] As Glaeser et al. (1992, p.1126) note, "After all, intellectual breakthroughs must cross hallways and streets more easily than oceans and continents."

[8] However, it is argued that geographic proximity is a necessary but not a sufficient condition for the transfer of tacit knowledge (Boschma, 2005; Torre and Rallet, 2005). Torre and Rallet (2005) outline four additional types of proximity that, in combination with geographic proximity, provide a sufficient condition for tacit knowledge transfer: cognitive proximity (a shared knowledge base), social proximity (socially embedded relationships), institutional proximity ('common habits, routines, established practices, rules, or laws') and organisational proximity ('the ability of an organisation to make its members interact'). (Also see Mokyr, 2005; D'Este et al., 2013; Torre and Rallet, 2005; Boschma, 2005; Hellmanzik, 2013; Audretsch, 1998; Rallet and Torre, 1998.) Yet, these types of proximity are difficult to quantify, and data on the individual or firm level is largely unavailable. Because of these limitations, research on learning-based agglomeration economies and knowledge spillovers has primarily relied on geographic proximity as a proxy for latent tacit knowledge transfer.

[9] This description is typically followed by reference to authors such as the Lake Poets (who are linked by having taken up residence in the rural Lake District in the early 19th century rather than by following a particular school of thought) or the Brönte Sisters (who spent most of their lives in the Yorkshire countryside in Northern England).

[10] While some authors may have enjoyed proximity to the readership due to fame, Romantic poets and Victorian novelists were typically anonymous, particularly for an author's first book. For example, Erickson (2002) estimates that by the "1770s sixty percent of all poetry books were published anonymously, but by the period's final decade from 1826 to 1835 only twenty-five percent were." A few authors, such as Edgeworth, Godwin and Opie, wrote with under their own name for later works, but the reading public generally did not demand to know or appear to care about authors' identities (St Clair, 2004). This is perhaps illustrated best by the fact that "throughout the Romantic period, all novels, whether anonymous or not, were normally classified in booksellers' lists and in circulating library catalogues, alphabetically by title, not under the name of the author" (St Clair, 2004, p. 174). The 'Author of Waverly' (Sir Walter Scott) is a notable exception as one of the few novelists that the reading public actively tried to identify. Jane Austen's identify, on the other hand, was only revealed years after her death.

need for a centralised location of production, the gains from agglomeration through these mechanisms may be so great that authors would still cluster geographically to avail of them. It is important to note that the greatest challenge to detecting spillovers is the 'Marshallian equivalence' of agglomeration mechanisms – that many different factors can result in the same observable outcome (Duranton and Puga, 2004; Carlino and Kerr, 2015). Therefore, it will not be possible to disentangle these mechanisms.

3. Data

3.1. Selection of authors

This paper utilises a unique, purpose-built dataset combining location and biographical information of prominent literary artists. This dataset was developed by collecting every individual associated with British or Celtic literature born between 1700–1925 with an entry in Encyclopaedia Britannica Online (2014).[11] To be defined as an author for the purposes of this study, an individual must have made at least one unique contribution to poetry or prose, which eliminated individuals whose contributions were strictly limited to translations, textbooks, manuals or guides, songwriting, literary criticism, or publishing.

Individual-level data was collected from three online encyclopaedias: Encyclopaedia Britannica (2014), The Literary Encyclopaedia (2014), and Literature Online (2014). These data include age, lifespan, age at first publication, number of publications per annum, lifetime publications, career duration, gender, and location for every year of the author's life.[12] Of the roster of 537 significant figures, only 370 have complete or near-complete lifetime location and publication information.[13] This sample of 370 authors is used in the econometric analysis in the following sections.

Although data were collected for the entire lifespan of each author, this study is limited to potential active years, defined as age 16 and older. This represents the lower bound for age at first publication. Additionally, many authors in the sample began university at age 17 or 18. Thus, potentially the first migration decision the authors made themselves would have been made around age 16 or 17. Because of this, the sample used in this analysis begins in 1725 (the first year with multiple authors aged 16+) and ends in 1999 (when the last authors died).

However, I am able to create a purpose-built index of author quality or impact defined as the amount of contemporary critical attention in academic literary studies given to one author relative to her peers, following the methodology of Murray (2003) and Verboord (2003). This index is constructed from two components: an indexed measure of the total words designated to an author from the three biographical sources and an indexed measure of the total number of citations within literary criticism as listed in the Literature Online (2014) database. These measures carry equal weight in the index. A sample of the top 15 literary artists and their respective Impact Index value is listed in Table 1 in B.2.

A more detailed description of the data collection process may be found in A.1. A more detailed description of the creation of the Impact Index may be found in A.2.

[11] Due to posthumous publishing and potential biases that may be associated with an increased interest and analysis in an author immediately following his or her death, living or recently deceased literary artists were not included. Thus, all literary artists in this dataset passed away before the year 2000.

[12] The publication data includes the total number of works published per annum with no quality adjustments.

[13] The term 'near-complete' is used specifically in regard to lifetime location because it was not unusual for individuals to have an unknown location for less than 5 years of their life. Unknown locations were particularly common during periods of conflict for males participating in military service. Such individuals were kept in the sample. With regard to publications, authors either had a comprehensive bibliography listed in Literature Online or a few select works listed across all three online dictionaries. Authors who do not have a comprehensive bibliographical entry were not included in the sample.

Table 1
Top 15 of authors by ranking.

Author	Word count index	Citation index	Impact index
Joyce, James	1	1	100
Dickens, Charles	0.75	0.76	75.47
Lawrence, D. H.	0.69	0.43	56.39
Conrad, Joseph	0.60	0.50	55.07
Woolf, Virginia	0.80	0.25	52.61
Blake, William	0.76	0.23	49.29
Wordsworth, William	0.56	0.42	48.86
Coleridge, Samuel Taylor	0.63	0.34	48.81
Austen, Jane	0.52	0.43	47.20
Eliot, T. S.	0.42	0.49	45.46
Yeats, William Butler	0.68	0.21	44.53
Byron, George Gordon	0.61	0.28	44.48
Beckett, Samuel	0.63	0.22	42.67
Johnson, Samuel	0.74	0.11	42.52
Hardy, Thomas	0.41	0.42	41.78

A note on interpreting these values: James Joyce has the most number of words in his biographical entries and the highest number of citations, so his Word Count Index (WCI) and Citation Index (CI) values are both equal to 1. The second-highest ranked author is Charles Dickens with a WCI value of 0.75, indicating that Dickens's biographical entries contained 75% the number of words as those of Joyce. Dickens' Impact Index value is 75.47 – the average of his WCI and CI values normalised to 100.

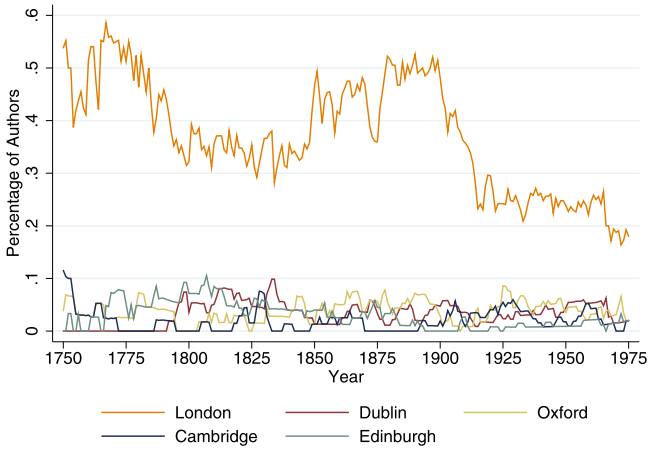

Fig. 1. Share ofauthors in major cities.

3.2. Trends in migration and clustering

The Greater London Area is the location with the most births of literary artists, with 79 (or 21.35%) of the artists born within this region. (See Table 2 in B.2.) I also evaluate spatio-temporal trends in lifetime movement to determine if authors tend to migrate and cluster in certain locations. As seen in Fig. 1, clustering intensity is surprisingly high throughout the sample. London consistently emerges as the largest geographic cluster of authors. Until the end of the 19th century, over 30% of all authors in a given year reside in London. (See A.1 for the definition of London over time.) A Herfindahl–Hirschman Index (HHI), defined as

$$HHI = \sum_{i=1}^{N} s_i^2$$

where s_i is the share of author-year observations of location i and N is the number of locations, can provide a more formal measure of author concentration. The HHI has a value of 0.17736, indicating a moderate degree of concentration.[14]

[14] For comparison, a market with a similar HHI value may be subject to scrutiny under US competition laws. See US Department of Justice and the Federal Trade Commission (2010) for further information.

Table 2
Number of authors born from 1700–1925, by birth location.

Region	1700–1724	1725–1749	1750–1774	1775–1799	1800–1824	1825–1849	1850–1874	1875–1899	1900–1925	Total per region
Connacht	0	0	0	1	0	0	3	0	1	5
East Midlands	1	2	1	1	4	3	0	3	1	16
East of England	2	1	4	0	6	5	0	3	3	24
Greater London	3	5	7	7	13	11	9	15	9	79
Leinster	0	2	1	6	2	1	4	4	4	24
Munster	1	0	0	1	1	1	0	0	2	6
North East England	1	0	0	0	1	0	0	1	0	3
North West England	0	0	1	3	3	1	3	2	3	16
Rest of Europe	0	2	0	1	0	1	5	1	1	11
Rest of World	0	0	0	0	1	2	4	6	5	18
Scotland	3	3	7	6	2	4	7	5	4	41
South East England	3	2	1	4	6	7	7	8	7	45
South West England	3	2	4	3	3	5	1	3	2	26
Ulster	0	0	0	0	0	0	1	4	2	7
Wales	0	1	0	0	0	1	3	0	5	10
West Midlands	2	1	0	0	2	1	3	6	6	21
Yorkshire and the Humber	0	0	0	0	3	1	1	4	5	14
Unknown	1	1	1	0	0	0	0	0	1	4
Total authors per period	20	22	27	33	47	44	51	65	61	370

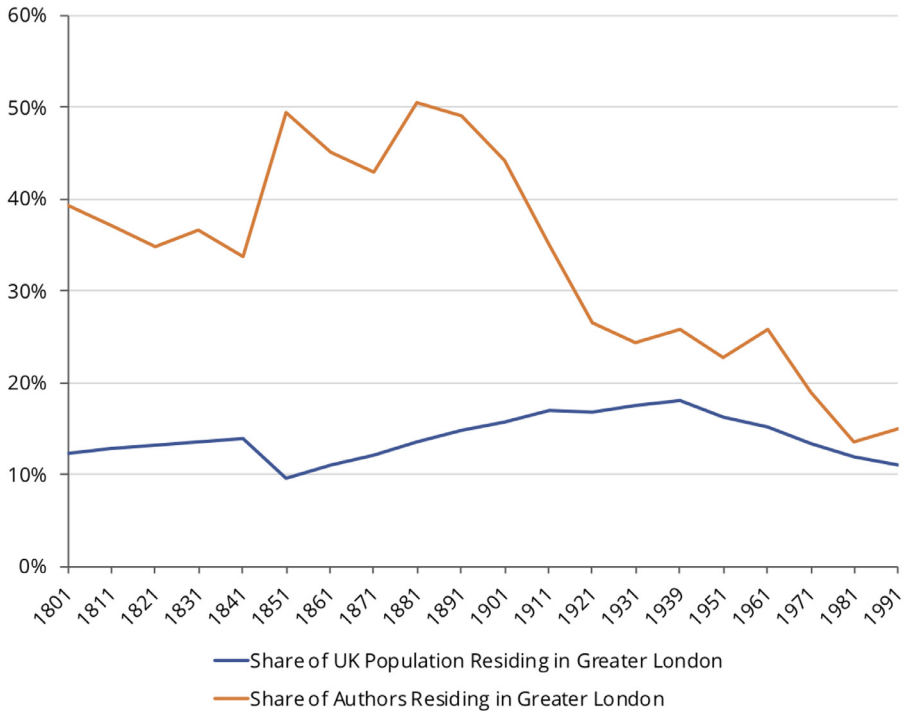

Fig. 2. Share of UK population and share of authors living in Greater London in census years 1801–1991.

The concentration of authors in London was greater than that of the total UK population, as seen in Fig. 2.[15] In each of the Census years from the first UK Census in 1801 until the last Census within the time frame of the sample in 1991, the share of authors living in Greater London exceeds the share of the total UK population living in Greater London. At the time of the 1851 Census (the approximate midpoint of the author sample), around half (49.3%) of the 75 authors lived in Greater London. In comparison, 9.7% of the 27.4m inhabitants of the UK resided in Greater London.

At its peak, London was home to over 50% of all authors. Dublin, Edinburgh, Oxford, and Cambridge emerge as the only other cities that

[15] Total population for Greater London are estimates of London's population derived from historic Census data. These figures are estimates to the nearest thousands (ONS calculation). These data were retrieved from the Office for National Statistics London Datastore (2014) website. The 1801–1841 total population data only includes Census data for England and Wales. These data were retrieved from the Great Britain Historical GIS / University of Portsmouth (2017) *A Vision of Britain through Time* website. The 1851–1991 total population data include mid-year estimates for the UK. These data were retrieved from the Office for National Statistics (2015) webpage for the ad hoc table "UK Population Estimates 1851 to 2014". The 1851–1901 total population data does not include the population of the Islands in the British Seas and of the portions of the Armed Forces and Merchant Service Abroad. There are no population figures available for Northern Ireland prior to 1911. The population figures prior to 1911 include data for Eire (Northern Ireland and the Irish Free State). There was no 1941 Census, therefore 1941 estimates are not available for the Greater London population. The only available figures are 1939 estimates. The 1911, 1939, and 1951 total population data includes civilians and home armed forces resident in the UK but does not include any UK armed forces posted from abroad. The 1921 and 1931 total population data includes civilians and all UK armed forces posted at home and abroad.

Table 3
Most populous cities (1851 census) vs top author destinations.

	England, Wales and Scotland*		Author destination	
Rank	City	Population	City	Author-year obs.
1	Liverpool	375,955	London	6620
2	Manchester	303,382	Oxford	777
3	Birmingham	232,841	Dublin	651
4	Leeds	172,270	Edinburgh	484
5	Glasgow	148,116	Cambridge	425
6	Bristol	137,328	Paris	265
7	Sheffield	135,310	New York	125
8	London+	127,869	Bath	105
9	Bradford (York)	103,778	Manchester	100
10	Newcastle upon Tyne	87,784	Norwich	84

Source: Population data is taken from the 1851 Census of Great Britain (Online Historical Population Reports, 2016). Data on top author destinations comes from author's calculations.

* Historical population data on Ireland are largely unavailable before the 1901 census, as only fragments remain of the 1821–1851 censuses. See National Archives of Ireland (2016) for more information.

+ London here refers to the borough of London within the *City* of London, as noted in the footnote of the original census table. For further discussion on the difficulties of defining 'London' over time, see Section 3.1.

see minor clustering of authors at any point in the sample. Due to the small sample sizes in these minor clusters (consistently fewer than 5 authors co-locating in each city per annum), I define London as the only cluster of authors for the primary analyses in this study. Due to the small sample sizes, I only utilise these cities in a robustness analysis to determine if locating in any one of the minor clusters results in an increase in productivity. (See Section 5.1.)

For comparison, Table 3 provides the 10 most populous cities in England, Scotland and Wales[16] in the 1851 census and the top 10 destinations for authors (in terms of author-year observations). Census data on city population is drawn from the Online Historical Population Reports (2016). While this only provides a single year observation for the most populous cities in England, there tends to be persistence in spatial distribution over time.[17] As can be observed in Table 3, authors did not tend to live in large British cities, with the exception of London. Manchester, for example, is the 9th most common author destination, yet 74 of the 100 author-years can be attributed to two authors (Samuel Bamford and Elizabeth Gaskell) who spent a substantial part of their lives in the city. All other authors lived in Manchester for less than 10 years. While there is some overlap years Bamford and Gaskell spent in Manchester, there is little to no overlap in the times other authors lived in the city. Therefore, it is not possible to disentangle a 'large city' effect from an 'intellectually diverse city' effect in this context of this paper.

It is notable that the clustering intensity of authors in London decreased from nearly 50% at the end of the 19th century to around 25% during the 20th century. (See Fig. 1.) This migration of authors out of London during this period was not accompanied by an increase in the clustering intensity in the minor literary clusters or a similar decrease in their clustering intensity. The decline of London in the early 20th century could be due to authors participating in or fleeing from the two World Wars. It could also be due to rise of Paris, New York and other cities as major destinations for authors. This may be an interesting and useful context to investigate the decline of one geographic cluster and the rise of another; although, this particular research question is beyond the scope of this paper.

I also investigate the variation in movement in and out of London, as a key concern is that London arises as a geographic cluster of authors only due to its role as a major location of birth. Of the sample of 370 authors, 71 authors never lived in London and only 4 authors never left London, indicating that approximately 80% of the sample spent at least some time living in London. Thus, the clustering intensity of London is not driven by the fact that London is a key birth location and is instead due migration in and out of London.

The trends in birth versus lifetime location reveal a surprising degree of mobility, migration, and co-location. The degree of clustering is significant in and of itself. There is no infrastructure required in the writing of books. Although there is infrastructure required in the production of books, little (if any) of the printing process required the author to be present. Transcripts could be mailed through the post, and many business dealings could be handled by a publisher or agent. "Felicia Hemans, for example, who lived in Wales, built her literary reputation first with Murray in London and later with Blackwood in Edinburgh, without visiting either city" (St Clair, 2004, p. 159).

3.3. Trends in productivity

As shown in Table 4, authors tended to be engaged in work-related activities for a great portion of their life: on average, authors lived for 66.5 years, published their first work at age 24 and had a career lasting just under 43 years. Over the course of their careers, authors produced an average of 33 publications, publishing an average of 0.79 works per year. Authors spent around 18 years of their adulthood (aged 16+) in London.

Although lifespans increased over time, the career trajectories of authors cohorts by century of birth are quite similar in terms of average career length and age at first publication. The cohorts born in the 19th and 20th centuries produced more works over the lifetime and had greater output per annum relative to the cohort born in 18th century. This is illustrated in more detail in Fig. 3, which includes violin plots of lifework works by birth cohort and a dashed red reference line to indicate the mean number of lifetime works for the entire sample of authors.[18] There is less variation in the number of lifetime works across authors for the 18th century cohort compared to the other birth cohorts. The 19th century cohort saw the first prolific authors with more than 100 works published over their lifetime, likely due to significant advances in

[16] Historical population data on Ireland are largely unavailable before the 1901 census, as only fragments remain of the 1821–1851 censuses. See National Archives of Ireland (2016) for more information.

[17] See Online Historical Population Reports (2016) for further additional census population reports.

[18] The box plot portion of the violin plot includes a point indicating the median, a box indicating the interquartile range, and 'whiskers' or 'spikes' illustrated the upper- and lower-adjacent values. The density plot illustrates the kernel density function.

Table 4
Summary statistics.

		Lifespan	Age at first publication	Career length	Lifetime works	Output per annum	Years in London
All authors	N = 370	66.52	23.66	42.86	33.21	0.79	17.89
		(16.22)	(11.37)	(18.74)	(32.23)	(0.66)	(18.41)
Century of birth							
18th century	N = 102	63.81	23.54	40.27	21.04	0.59	19.59
		(16.95)	(11.36)	(17.95)	(17.74)	(0.51)	(18.99)
19th century	N = 207	67.00	23.13	43.87	39.57	0.89	18.45
		(16.04)	(11.73)	(19.25)	(37.50)	(0.73)	(18.38)
20th century	N = 61	69.46	25.69	43.77	31.97	0.77	13.16
		(15.17)	(10.01)	(18.15)	(25.23)	(0.53)	(16.96)
Impact index quartile							
Lowest quartile	N = 91	66.45	28.24	38.21	18.47	0.53	17.36
		(16.27)	(12.25)	(19.10)	(21.21)	(0.60)	(20.44)
Second quartile	N = 94	66.78	22.93	43.85	33.81	0.78	18.94
		(16.36)	(11.27)	(16.66)	(29.87)	(0.62)	(17.29)
Third quartile	N = 93	66.09	21.76	44.32	32.90	0.78	18.80
		(15.29)	(10.77)	(18.34)	(33.22)	(0.64)	(19.14)
Top quartile	N=92	66.78	21.80	44.98	47.49	1.07	16.43
		(17.18)	(10.00)	(20.25)	(36.26)	(0.66)	(16.74)

This table reports the mean of each variable with the respective standard deviation reported in parentheses.
Because many authors were publishing throughout their entire lives, career length is defined as Lifespan - Age at First Publication. Output per annum is thus defined as Lifetime Works / Career Length.
The number of years in London only include the number of years authors were potentially actively engaged in their career (defined as ages 16+ in Section 3.1) spent in London and not the total number of years from birth.

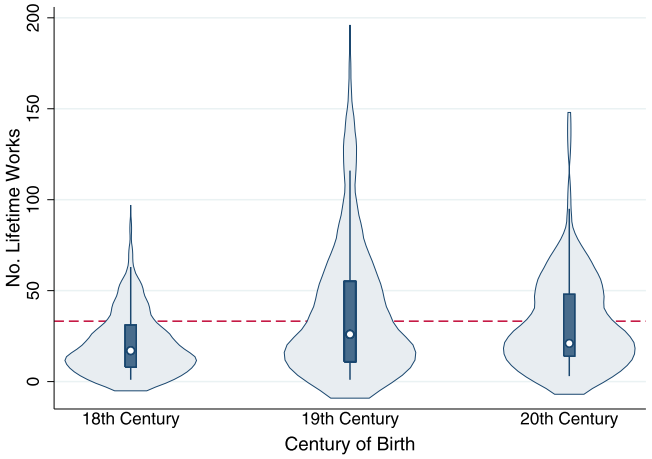

Fig. 3. Distribution of lifetime works by century of birth.

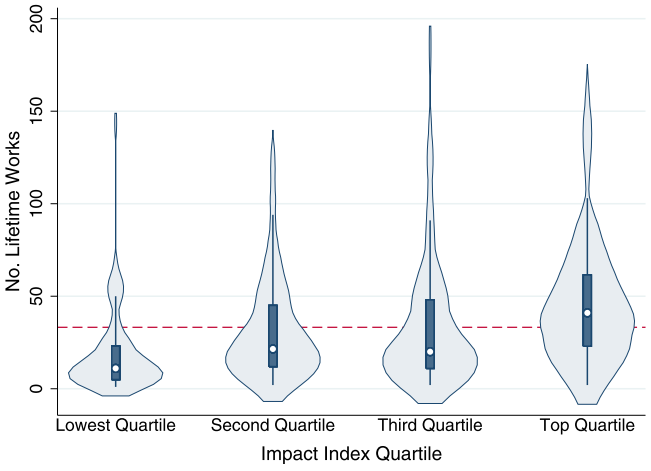

Fig. 4. Distribution oflifetime works by impact index quartile.

printing technology (such as the steam-powered printing press) during this century. The decline of London as a cluster is also reflected in the amount of time authors spend living there: the average author born in the 18th century spent around 20 years in London; whereas, the cohort born in the 20th century spent only 13 years in London.

There are also differences in career trajectories by Impact Index quartile. The lifespans of these 'quality' or 'prominence' cohorts does not differ substantially, and all cohorts spent an average of 16–18 of their adult years in London. However, the cohort of the least prominent authors in the sample (those ranking in the lowest quartile) began publishing around age 28, much later than the other cohorts who began publishing around age 22. This group also produced, on average, far fewer works over their lifetimes compared to the other cohorts: 18 works for the lowest quartile compared to 33 works for the middle quartiles and 47 works for the top quartile.[19]

The distribution of lifetime works by Impact Index quartile is illustrated in more detail in the violin plots in Fig. 4. Across all cohorts, the largest concentration of authors produce less than 50 works over their lifetimes. Authors in the lowest quartile tended to publish relatively few works over their lifetimes compared to the other cohorts; although, there is at least one prolific author who produced just under 150 over the lifetime. The distributions for the cohorts of authors ranking in the second and third quartiles are quite similar, although there are some exceptionally prolific authors in the third quartile cohort. The cohort of the most prominent authors is more normally distributed with a higher peak around 40 works and more prolific authors.

Fig. 5 shows the total number of authors and total output (of all authors) per annum over time. There is no significant change in total production for the first 150 years, followed by a sharp and steady

[19] Although there appears to be a positive relationship between the cohort means for lifetime works and the Impact Index quartile, there is only a weak correlation between the number of lifetime works and Impact Index score (r = 0.2972).

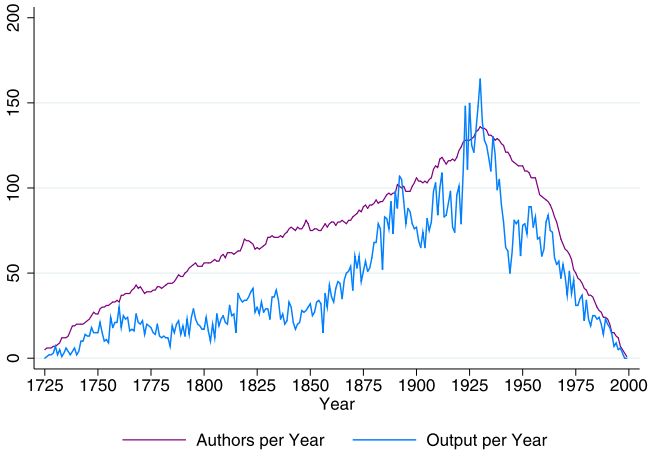

Fig. 5. Total authors and output per annum.

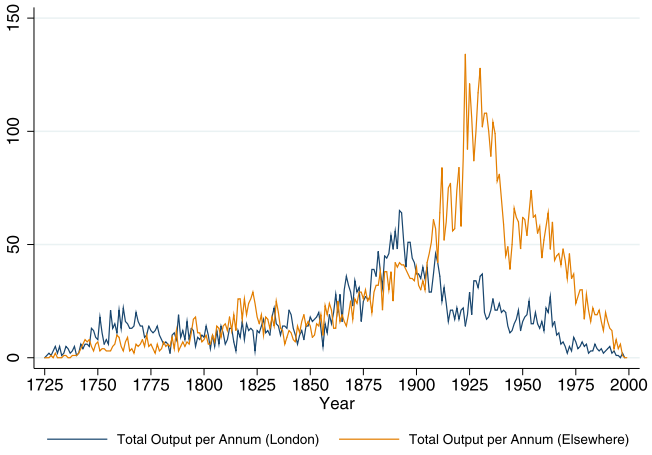

Fig. 6. Total output per annum by location.

increase in total production from around 1850. It is notable that this is not accompanied by a corresponding dramatic increase in the total number of authors.[20] The dynamics of publishing are further detailed in Fig. 6, which decomposes the total number of publications between those published by authors in London and those published by authors living in all other locations. Until the turn of the twentieth century, authors residing in London consistently produce at least as many publications as all authors in all other locations combined.

4. Identification strategy

I begin by determining if living in a geographic cluster (London) results in an increase in productivity and estimating these returns. A standard approach is to estimate the following relationship by ordinary least squares using a pooled panel of workers:

$$Output_{it} = \delta Location_{it} + \boldsymbol{\beta}\mathbf{x}_{it} + \upsilon_{it} \quad (1)$$

in which some measure of productivity (outut per annum) is a function of the location in which worker i resides in time t and a vector of individual-specific characteristics.

An OLS estimation of Eq. (1) will yield unbiased and consistent results only if $\mathbf{x}_{it}$ is measured without error; however, some elements of $\mathbf{x}_{it}$ are unobservable. Specifically, Eq. (1) suffers from potential source of bias, as it does not account for spatial sorting by worker skill. There is an endogenous relationship between productivity and location, as there is likely to be some unobserved individual-level characteristic, such as natural ability, that is linked to an individual's likelihood of migrating to a particular location. If individuals with higher unobserved ability are more likely to migrate, then the OLS estimate of $Location_{it}$ will be upwardly biased. The bias will be downwards in the opposite case.

A standard solution is to find an exogenous instrument for location choice and then estimate the pooled OLS.[21,22] However, if panel data on individual workers is available (as is the case in this paper), (Glaeser and Maré, 2001; Combes et al., 2008; de la Roca and Puga, 2012) argue that the use of individual fixed-effects can adequately address the issue of workers sorting across locations on unobservables.[23]

I first provide OLS estimates for the pooled panel of authors as a baseline estimate of the correlation between location and productivity along the cross-section. I then address the identification issue discussed above by estimating the following equation:

$$output_{it} = \delta London_{it} + \beta_1 age_{it} + \beta_2 age_{it}^2 + \sigma_i + \mu_t + \epsilon_{it} \quad (2)$$

where $output_{it}$ measures the number of known works by author i published in time t.[24] The variable of interest is $London_{it}$ is a binary variable equal to 1 if author i was living in London in time t. The variables age_{it} and age_{it}^2 control for author-specific life-cycle ageing effects, with the quadratic term allowing for productivity to diminish as a author's age increases. μ_t is a vector of time dummies to account for yearly changes in productivity that impact all authors in the same way. I also include a full set of author fixed effects, given by σ_i. The combination of the linear component of the age profile, 1-year dummy variables and individual author fixed effects can lead to the "age-time-cohort" identification problem, so this estimation is repeated with 2-year and 5-year time dummies.

Furthermore, there are many years in which authors do not publish at all. In years in which they do publish, most authors publish only a single work per year, with increasingly fewer having two, three, etc. publications per year, as seen in Fig. 7. Therefore, the study is limited to potential active years (age 16 and older), as defined in Section 3.3. I also supplement the analyses with a negative binomial

[20] This shift in total production could be due to invention and widespread of industrial printing press during this time or the continual increase in the total number of authors led to an increased competitiveness that manifested itself in increased total output. Perhaps, as well, this is due to rise in rise in literacy and the advent of the "penny dreadful" and "shilling shocker" in the mid to late 19th century. Most likely, it is some combination of all these factors (St Clair, 2004).

[21] Ciccone and Hall (1996) suggest using long lags of population density to instrument for the size or density of local population. They argue that the spatial distribution of the population persists over time and that this differs substantially from the contemporary factors impacting productivity on the local level.

[22] Borowiecki (2013) uses birthplace-cluster distance to instrument for the spatial sorting of prominent Western classical music composers, as birthplace to cluster distance captures not only the costs associated with travel but also factors such as cultural and linguistic distance. Since most specifications define London as the only cluster, it is not possible to use a time-invariant instrument and individual fixed effects. Additional detailed individual-level information would be needed to control for individual heterogeneity in order to utilise birthplace-London distance as an instrument, which is often not available in the limited historical sources. Furthermore, given the limited geographic scope of this paper, the travel distances are much smaller and the culture and language less varied. Particularly after the development of extensive rail lines by the mid-1800s, distance across the UK was not likely such a deterring factor in moving to London.

[23] Combes and Gobillon (2015) note that "as long as individual location decisions depend only on the explanatory terms introduced in the specification, which can go as far as the individual fixed effect, some time-varying individual characteristics such as age, and a location-time fixed effect, there is no endogeneity bias."

[24] I rely on observable measures of productivity for our analysis. As mentioned in the previous section, productivity is measured in terms of number of known publications rather than number of works written. I have no measure of the quality of individual publications.

Table 5
Main results.

	(1) OLS	(2) Fixed effects	(3) FE with 2-year dummies	(4) FE with 5-year dummies	(5) Negative Binomial IRR
Age	0.0766***	0.0870***	0.0762***	0.0757***	1.1603***
	(0.00511)	(0.00549)	(0.00474)	(0.00592)	(.00542)
Age-squared	−0.000728***	−0.000738***	−0.000712***	−0.000735***	0.99843 ***
	(0.0000510)	(0.0000492)	(0.0000473)	(0.0000487)	(0.0000425)
London	0.0331	0.0944***	0.0806**	0.0939***	1.2435***
	(0.0528)	(0.0364)	(0.0374)	(0.0362)	(0.04003)
Constant	−1.232***	−1.562***	−1.120***	−0.360	0.1173***
	(0.134)	(0.158)	(0.106)	(0.643)	(0.0312)
Time dummies	1-Year	1-Year	2-Year	5-Year	1-Year
Author FE	No	Yes	Yes	Yes	Yes
R^2	0.103	0.120	0.096	0.111	
No. authors	370	370	370	370	370
Observations	19,022	19,022	19,022	19,022	19,022

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Robust standard errors are clustered on the author level and are reported in parentheses.

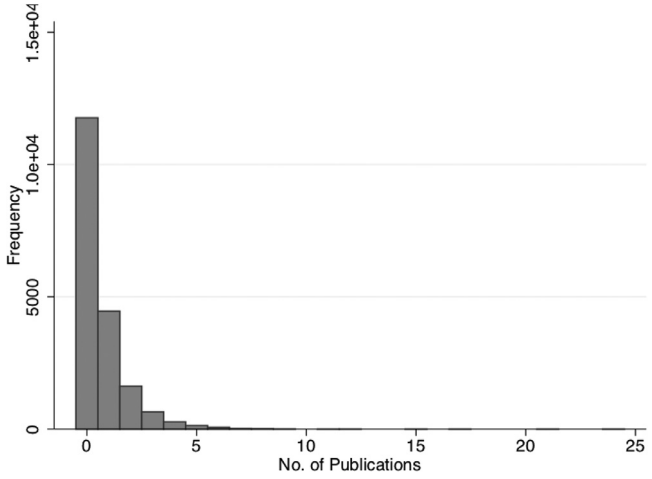

Fig. 7. Frequency of output observations.

model, which is a more appropriate specification for over-dispersed count data (Wooldridge, 2010).[25]

However, establishing a causal link is a challenge due to the potential endogeneity of the propensity to cluster. If the only source of endogeneity is spatial sorting by skill level, this issue is addressed through inclusion of author fixed effects controls. If, however, there is a dynamic feedback effect between past output and the probability of migrating to London (i.e. authors with recent career success systematically migrate to London at higher or lower rates and that career success is not linked to innate ability), the fixed effects estimates will be biased. This issue is addressed in greater detail in Section 5.3.

It is also important to note that, due to the nature of the data, it is not possible to identify the channels through which authors receive gains or the degree to which these mechanisms are geographically localised. It is also not possible to determine whether any effects at work are due to the size of the city (urbanisation) or the concentration of the industry (localisation).

5. Results

The estimates for Eq. (2) are presented in Table 5. Column (1) shows the estimates of the OLS relationship between locating in the London cluster and the number of publications in a given year using the pooled panel of authors, utilising cross-sectional variation to identify the effect of living in London. In Column (2), I introduce author fixed effects, drawing on temporal variation within individuals to identify the effect of living in London while controlling for sorting by skill level. Columns (3) and (4) provide the results with 2-year and 5-year time dummies, respectively. These results are consistent with the estimates in Column (2), although the effect is more modest. Column (5) includes the incident rate ratios for the negative binomial model with individual fixed effects. In all five models, robust standard errors are clustered on the individual level to account for serial correlation within individuals.

As seen in Column (1), much of the variation in production between individuals cannot be explained by location. However, when individual fixed effects are introduced in Column (2), an author residing in London experiences productivity gains of 0.094 additional works per annum compared years of her life when residing elsewhere. Given that the mean output per annum is 0.79, this translates to a 11.94% increase in annual productivity. The negative binomial incident rate ratios (NB IRR) reported in Column (5) indicate that an author sees a 24.3% increase in the probability of publishing in a given year while residing in London compared to years of her life when residing elsewhere.

In all specifications, age has a positive but diminishing effect, which is consistent with general findings in the literature on individual productivity.[26] The R-squared value of the fixed effects model is also notable: 12% of the variation in productivity can be explained by only four variables (age, London, and time and author fixed effects). Creative production is often anecdotally viewed as a highly ethereal process – one that is organic and impalpable, beyond quantification. Yet, a substantial amount of literary production can be explained by a simple life-cycle production process and geographic proximity, not so dissimilar to the production processes of R&D researchers. (See Levin and Stephan, 1991.)

However, a key concern is that the results may simply reflect the fact that authors are positively selected into London. That is, it is possible that the "London effect" reflects only the fact that the most productive authors are the most likely to migrate to London. In this respect, the difference in the estimates in Column (1) and Column (2) in Table 5 is important. The results of the OLS regression indicate that location is not correlated with productivity in the cross-section – i.e., there is no statistically significant relationship between author-year observations in London and high levels of output. Therefore, it is not the case that only the authors who published the greatest number of works in a given year lived in London. However, the results of the regression with author fixed effects presented in Column (2) indicate that authors were more productive during the periods of their lives when they resided in London

[25] Although not reported, the conditional variance of the output variable is greater than the conditional mean; therefore, the data is over-dispersed.

[26] See, for example, Levin and Stephan (1991).

Table 6
Effect of minor clusters.

	OLS fixed effects		Negative binomial IRR	
	Minor clusters	All clusters	Minor clusters	All clusters
Age	0.0761***	0.0865***	1.161***	1.160***
	(0.00499)	(0.00546)	(0.00540)	(0.00544)
Age-squared	−0.000725***	−0.000735***	0.998***	0.998***
	(0.0000501)	(0.0000487)	(0.0000427)	(0.0000427)
Minor clusters	−0.0969*	−0.0433	0.925	0.985
	(0.0511)	(0.0420)	(0.0466)	(0.0507)
London		0.0881**		1.241***
		(0.0371)		(0.0407)
Constant	−1.195***	−1.546***	0.128***	0.118***
	(0.133)	(0.158)	(0.034)	(0.031)
Year dummies	Yes	Yes	Yes	Yes
Author FE	Yes	Yes	Yes	Yes
R^2	0.103	0.120		
No. authors	370	370	370	370
Observations	19,022	19,022	19,022	19,022

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Robust standard errors are clustered on the author level and are reported in parentheses. The heading 'Negative Binomial IRR' denotes that the incident rate ratios are reported for the negative binomial regressions.
The 'Minor Clusters' is a binary variable equal to 1 if the author lived in Cambridge or Dublin or Edinburgh or Oxford in a given year and equal to 0 otherwise.

compared to when they lived elsewhere. Whether authors also receive gains from living in other locations is examined in the following section.

5.1. Agglomeration effects in other cities

It is possible that authors also receive productivity gains from locating in other cities. Authors may receive a 'large city' effect from locating in one of the most populous cities. Particularly for skilled workers, such cities would offer more alternatives if a literary career was unsuccessful or if the author needed to supplement their income from writing with other sources. Authors may also receive productivity gains from locating in Edinburgh, in particular, due to the well-developed local book publishing infrastructure and low costs of production. Authors may benefit from locating in smaller university cities, in which exposure to a diverse intellectual elite promotes an increase in creative activity. The primary difficulty in examining any of these potential factors is the low degree of concentration in non-London cities. As discussed in Section 3.2, authors did not tend to locate in the most populous cities in the United Kingdom. The only non-London cities that are consistently a location to multiple authors in the same year are Cambridge, Dublin, Edinburgh and Oxford.

Because there are so few observations for the minor cluster cities (Cambridge, Dublin, Edinburgh and Oxford), it is not possible to evaluate the effect of each city separately within the framework utilised in this paper. It is possible to determine the effect of an author locating in any one of the minor clusters by defining a general 'minor clusters' indicator: a binary variable equal to 1 if the author lived in Cambridge or Dublin or Edinburgh or Oxford in a given year and equal to 0 otherwise. Even at this level of aggregation, 48.58% of the sample of authors have a standard deviation of 0, indicating they either spent no time in any of the minor clusters or they spent their entire lives in the clusters. Given that so few authors were present in any of these cities in a given time (as discussed in Section 3.2), it can be assumed most of these authors never lived in any of the minor clusters.

I follow the same strategy described in Section 4 and present the results in Table 6. Columns 1 and 2 provide the coefficients for an OLS fixed effects estimation, and Columns 3 and 4 present negative binomial fixed effects incident rate ratios. Columns 1 and 3 only include the 'Minor Clusters' indicator, while Columns 2 and 4 also include the indicator for residing in London. These results should be interpreted cautiously, as there is not much within-author variation for the 'Minor Clusters' variable. Locating in any one of the minor clusters has a negative and significant impact on productivity in the OLS fixed effects specification but does not have a significant effect when controlling for periods when residing in London. In the negative binomial models, locating in a minor cluster has no significant impact on productivity, although the coefficient also indicates a negative relationship.

In general, these results suggest that there is something unique about the conditions of London that is not present in other literary cities, including the rival publishing city of Edinburgh. Access to literary infrastructure alone is not enough for authors to receive an increase in productivity. Clustering intensity could be one factor. It could be that there is a minimum number of authors needed in a location for gains to realise. Productivity is likely to increase as the number of peers increases due, for example, to an increase in local competition, stronger local social networks, a reduction in how costly it is for a new innovation to be realised or increased knowledge spillovers. On the other hand, the diversity of industry or concentration of other creative industries in London could also be important features. Without detailed data, these potential mechanisms cannot be empirically tested.

5.2. Heterogeneity of returns

There is substantial heterogeneity of returns by century of birth and by Impact Index quartile, as seen in Table 7. In general, these findings of heterogeneity of returns across birth cohorts and Impact Index quartiles indicate that the observed "London effect" is not a result of authors in London needed to produce at a higher and steady rate to cover the high cost of living in London. These results also reveal insight into the potential mechanisms underlying the observed London premium and provide some assurances that this finding is not driven by specific cohorts of particularly prolific authors.

Authors born in the 18th century receive the substantial productivity gains from locating in London, producing 0.191 additional works per year while living in London relative to the years when living elsewhere. This effect is not seen for the cohorts of authors born in the 19th or 20th centuries. This is an important finding. The cohort of authors born in the 18th century produced fewer works over their lifetime compared to the later birth cohorts, and exceptionally prolific authors (publishing 100+ works over their lifetime) are only present in later birth cohorts. Therefore, the "London effect" here is not being driven by the most prolific authors. That this effect is not observed for later birth cohorts could be a result of improved printing, transportation and communication during the 19th and 20th centuries, which would allow authors to interact with frequency and relative ease without needing to

Table 7
The 'London Effect' by century of birth and impact index quartile.

	Century of birth			Impact index quartile			
	18th century	19th century	20th century	Lowest quartile	Second quartile	Third quartile	Top quartile
Age	−0.0181	0.0894***	0.0872***	0.0201	0.0961***	0.0833***	0.107***
	(0.188)	(0.00803)	(0.0163)	(0.114)	(0.0132)	(0.00964)	(0.0113)
Age-squared	−0.000517***	−0.000854***	−0.000729***	−0.000455***	−0.000842***	−0.000750***	−0.000969***
	(0.0000687)	(0.0000802)	(0.000151)	(0.0000794)	(0.000126)	(0.0000936)	(0.000105)
London	0.191**	0.0460	0.0828	0.0684	0.00360	0.175**	0.174***
	(0.0763)	(0.0496)	(0.0750)	(0.0776)	(0.0669)	(0.0779)	(0.0632)
Constant	−3.166	−0.903***	−1.008***	−0.0595	−2.186***	−1.728***	−1.704***
	(5.611)	(0.158)	(0.253)	(2.480)	(0.353)	(0.302)	(0.365)
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Author FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.077	0.140	0.127	0.099	0.179	0.161	0.199
No. authors	102	207	61	91	94	93	92
Observations	4967	10,756	3292	4683	4867	4740	4732

Robust standard errors are clustered on the author level and are reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

living within London. This also could be due, in part, to decline in the concentration of authors in London and the rise of other major cultural centres in cities such as Paris and New York in the 20th century.

There is also evidence that the returns from agglomeration are limited to the most prominent authors (those who rank in the second and top Impact Index quartiles). The difference between the cohorts rankings in the second and third quartiles is striking. These cohorts produced, on average, a similar number of works over the lifetime, had similar rates of output per year, and spent a similar number of years in London. Yet, only the more prominent cohort of authors received a premium from living in London. It could be that these authors had access to better local social networks (both within literary circles and across other creative industries), potentially providing both information on local publishers and literary critics and preferential access to inputs through social ties. Thus, on average, the cohort of authors ranking in the second quartile were more productive during the periods of their lives in which they were in London, where as the cohort ranking in the third quartile produced at similar rates regardless of whether they lived in London.

5.3. Spatial selection as a dynamic process

Empirically, establishing a causal link is a challenge due to the potential endogeneity of the propensity to cluster. This is controlled for by the inclusion of author fixed effects if the only source of endogeneity is spatial sorting by skill level; however, it is still possible that there is a dynamic feedback effect between past output and the probability of migrating to London. In this case, the strict exogeneity assumption is violated, and the estimates in Table 5 will be biased. The primary concern in this context is that authors' recent career success – who are not necessarily those with greater innate ability – systematically migrate to London at higher rates. In this case, the fixed effects estimates will be overestimates of the true effect. Similarly, if authors who experience recent career difficulty move to London at systematically higher rates – potentially believing this will improve their chances of experiencing a 'big break' or will provide them with more career alternatives if their pursuit of a literary career is unsuccessful – the fixed effects estimates will be a lower bound.

I explicitly address this issue by using a linear probability model to determine the role of past success (in terms of output) in the probability of migrating to London. Specifically, I estimate the following regression:

$$migrate_{it} = \gamma_1 OutputToDate_{it} + \beta_1 age_{it} + \beta_2 age_{it}^2 + \sigma_i + \mu_t + \epsilon_{it} \quad (3)$$

in which $migrate_{it}$ is equal to 1 in the year author i migrates to London and equal to 0 otherwise, $OutputToDate_{it}$ is the cumulative number of works that an author i has published from the beginning of her career until time t. This equation is first estimated for the entire sample of authors and then estimated by century of birth and by Impact Index quartile to determine if the results are being driven by a particular cohort.

The results are presented in Table 8. Column (1) provides the relationship between total output to date and the probability of migrating to London for the entire sample of authors. Columns (2)–(4) presents the results by century of birth: authors born in the 18th century, 19th century, and 20th century, respectively. The final four columns, Columns (5)–(8), present the results by impact index quartile, beginning with the authors in the lowest quartile in Column (5) and ending with the authors in the top quartile in Column (8).

For the results of the specification with all authors, the coefficient for total output to date is negative, suggesting negative selection to London (i.e. authors who had more publications were less likely to migrate). However, there is not a statistically significant relationship between total output to date and migration to London. The result with respect to the negative selection of authors is generally consistent across all specifications. A notable exception is the cohort of authors who rank in the lowest quartile of the Impact Index, where there is potentially evidence of positive selection among this cohort of authors. However, this relationship is not statistically significant.

The only statistically significant relationship observed is for the cohort of authors born in the 18th century. The results for this cohort of authors suggests negative selection. This is an important finding, given that this is the birth cohort that receives the greatest gains from the agglomeration of literary activity in London. For the cohorts of authors ranking in the top two quartiles of the Impact Index (who also received substantial gains from living in London), there no evidence of positive selection. However, there is no evidence clearly indicating negative selection either.

As an additional robustness check, I estimate the same set of specifications using output in the previous period instead of the total output to date. These results are presented in Table 9. There is no statistically significant relationship between output and migration in any of the specifications; however, the coefficient for output in the previous period is positive in several specifications.[27]

5.4. Career trajectories before and after migration to London

In order to examine the London premium in more detail, I examine productivity in the years leading up to and following migration to London, focusing only on the authors who migrate (thus excluding

[27] Similar regressions using a logit model instead of a linear probability model (not reported) yield findings that are consistent with the results for the linear probability model; however, I chose to report the linear probability models as the standard errors are more directly interpretable.

Table 8
Probability of migrating to London: Total output to date.

	All	Century of birth			Impact index quartile			
	Authors	18th century	19th century	20th century	Lowest quartile	Second quartile	Third quartile	Top quartile
Age	−0.00271***	−0.00228	−0.000973**	−0.00208	0.0458	−0.00162**	0.00351	−0.00270***
	(0.000336)	(0.0184)	(0.000477)	(0.00144)	(0.0467)	(0.000813)	(0.00274)	(0.000693)
Age-squared	0.00000620*	−0.00000229	0.00000745	0.00000238	0.00000592	0.0000115	−0.00000378	0.00000707
	(0.00000340)	(0.00000693)	(0.00000478)	(0.0000137)	(0.00000593)	(0.00000717)	(0.00000711)	(0.00000772)
Output to date	−0.000181	−0.000802***	−0.00000834	−0.000312	0.0000601	−0.0000795	−0.000365	−0.000250
	(0.000123)	(0.000285)	(0.000139)	(0.000422)	(0.000206)	(0.000243)	(0.000306)	(0.000272)
Constant	0.105***	−0.0886	0.0190***	0.0244	−0.868	0.0462***	−0.0310	0.0364**
	(0.0272)	(0.550)	(0.00559)	(0.0200)	(1.033)	(0.0167)	(0.0856)	(0.0175)
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.022	0.039	0.024	0.031	0.068	0.070	0.066	0.074
No. authors	370	102	207	61	91	94	93	92
Observations	18,730	4967	10,480	3281	4671	4795	4667	4597

Robust standard errors are clustered on the author level and are reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 9
Probability of migrating to London: Output inprevious period.

	All	Century of birth			Impactindex quartile			
	Authors	18th century	19th century	20th century	Lowest quartile	Second quartile	Third quartile	Top quartile
Age	−0.00338***	−0.00604	−0.00252***	−0.00322**	0.0445	−0.00272***	0.00276	−0.00343***
	(0.000388)	(0.0194)	(0.000914)	(0.00155)	(0.0468)	(0.000814)	(0.00265)	(0.000783)
Age-squared	0.0000118***	0.00000413	0.0000144***	0.0000104	0.0000108	0.0000204***	0.000000597	0.0000135
	(0.00000370)	(0.00000782)	(0.00000520)	(0.0000145)	(0.00000676)	(0.00000757)	(0.00000750)	(0.00000816)
Output in previous year	−0.00000410	−0.00335	0.000464	0.00250	−0.000581	0.00384	−0.00409	0.000955
	(0.00119)	(0.00321)	(0.00135)	(0.00327)	(0.00193)	(0.00249)	(0.00252)	(0.00230)
Constant	0.124***	0.652	0.134**	0.218***	−0.826	0.0698***	−0.00396	0.0592***
	(0.0284)	(2.436)	(0.0561)	(0.0380)	(1.033)	(0.0163)	(0.0840)	(0.0196)
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.024	0.041	0.025	0.032	0.070	0.074	0.068	0.075
No. authors	370	102	207	61	91	94	93	92
Observations	18,652	4865	10,549	3231	4592	4773	4647	4640

Robust standard errors are clustered on the author level and are reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

those who always lived in London or never lived in London). This provides a way to determine whether the authors who migrated to London were already experiencing a build-up of successful publishing and simply continued to be successful after their arrival. If such a trend exists, then the estimated London premium would reflect the continuation of an existing career development process rather than gains explicitly associated with agglomeration of literary activity.

It is possible that authors migrated in and out of London multiple times within their lifespan. It is likely that the greatest gains are observed in the first instance of migration, as this is when authors first receive access to shared inputs, develop high-quality matches to key people in local networks, and learn 'secrets of the trade'. Although an author may leave London at some point, the knowledge learned is retained and social connections may be easily maintained. If an author returns to London again, there may be some additional gains but, arguably, these will not be as substantial as with the first instance of migration. Therefore, I only evaluate the trend in productivity pre- and post- migration to London for the first instance of migration.

Specifically, I first estimate the following equation:

$$output_{it} = \gamma_1 \mathbf{PreMigration_{it}} + \gamma_2 \mathbf{PostMigration_{it}} + \beta_1 age_{it} + \beta_2 age_{it}^2 + \sigma_i + \mu_t + \epsilon_{it} \quad (4)$$

in which $\mathbf{PreMigration_{it}}$ is a set of dummy variables for each of the 5 years before individual i migrates to London and $\mathbf{PostMigration_{it}}$ is a set of dummy variables for each of the 5 years after individual i migrates to London. The $\mathbf{PostMigration_{it}}$ dummies are only equal to one if the author is living in London, thus excluding observations where an author migrates out of London. These sets of pre- and post-migration dummies also include one dummy variable for all the years before 5 years before migration and one for all years after 5 years after migration. As with the previous regression, this exercise is repeated for all authors, for birth cohorts (century of birth), and for Impact Index cohort (Impact Index quartiles). The coefficients for the pre- and post-migration dummies for all specifications are reported in Table 10.

The coefficients for the pre- and post-migration dummies using the entire sample of authors are plotted in Fig. 8, along with the respective 95% confidence intervals. In this graph, the coefficient for $t+2$ is interpreted as the productivity of individual i in the second consecutive year in London after migrating relative to the year in which she first migrated to London. For the five years before the year of migration, the productivity is not statistically different from the year of migration.[28] However, there is a statistically significant increase in productivity in the first three years after migration to London. Productivity in the fourth year after migration is not statistically different from that in the year of migration; although, there is a statistically significant increase in productivity from the following year. These results suggest that the gains from living in London accumulate over time, which is consistent with the dynamic externalities associated with living in a city observed by Glaeser and Maré (2001) and others.

The estimated coefficients for the birth cohorts are illustrated in Fig. 9. The productivity of all birth cohorts before migration is not

[28] Although there appears to be an increase in productivity in the year before the author migrates to London, the 95% confidence interval crosses zero. Therefore, there is no statistically significant difference between productivity in this period and productivity in the period in which the author migrates to London.

Table 10
Estimated trend pre- and post-migration.

	(1) All authors	(2) 18th century	(3) 19th century	(4) 20th century	(5) Lowest quartile	(6) Second quartile	(7) Third quartile	(8) Top quartile
t< t−5	−0.00967	0.0311	0.0892	−0.0955	−0.0749	0.0986	0.0131	−0.0588
	(0.0811)	(0.134)	(0.0836)	(0.0863)	(0.120)	(0.115)	(0.109)	(0.134)
t−5	−0.00927	−0.0592	−0.0166	−0.0182	−0.116	−0.0303	0.0682	−0.132
	(0.0734)	(0.121)	(0.0828)	(0.140)	(0.0951)	(0.131)	(0.128)	(0.127)
t−4	0.0314	−0.159	0.104	−0.0586	0.0513	0.0510	−0.0672	−0.0833
	(0.0598)	(0.107)	(0.0765)	(0.0890)	(0.0914)	(0.0997)	(0.110)	(0.140)
t−3	−0.00154	0.0346	0.0330	−0.0230	−0.0260	−0.0255	−0.00678	0.0580
	(0.0574)	(0.120)	(0.0698)	(0.117)	(0.0964)	(0.0963)	(0.114)	(0.113)
t−2	0.0398	−0.0569	0.0185	0.143	−0.0161	0.0909	0.0356	−0.0485
	(0.0557)	(0.121)	(0.0570)	(0.0986)	(0.0879)	(0.0797)	(0.110)	(0.102)
t−1	0.0960	−0.0477	0.100	0.111	−0.0905	0.255***	−0.0741	0.0541
	(0.0593)	(0.0956)	(0.0753)	(0.101)	(0.0778)	(0.0906)	(0.0810)	(0.133)
t+1	0.127**	0.207*	0.0765	0.00418	−0.0662	0.0116	0.230**	0.102
	(0.0559)	(0.120)	(0.0700)	(0.0942)	(0.0657)	(0.0791)	(0.103)	(0.143)
t+2	0.191***	0.311**	0.145**	0.0501	0.0818	0.0180	0.329**	0.105
	(0.0591)	(0.119)	(0.0686)	(0.110)	(0.0960)	(0.0750)	(0.129)	(0.120)
t+3	0.183***	0.239*	0.0952	0.0809	−0.00673	−0.0116	0.244**	0.189
	(0.0656)	(0.133)	(0.0669)	(0.124)	(0.108)	(0.0819)	(0.115)	(0.128)
t+4	0.0815	0.121	−0.0352	0.119	−0.193	0.0650	0.132	0.0244
	(0.0653)	(0.115)	(0.0741)	(0.153)	(0.127)	(0.115)	(0.122)	(0.123)
t+5	0.280***	0.230	0.164*	0.132	−0.0655	0.189	0.340**	0.167
	(0.0882)	(0.151)	(0.0952)	(0.141)	(0.158)	(0.122)	(0.146)	(0.177)
t> t+5	0.193**	0.170	0.130	0.111	−0.155	0.172	0.257**	0.285**
	(0.0919)	(0.103)	(0.0876)	(0.117)	(0.133)	(0.119)	(0.104)	(0.124)
R^2	0.145	0.104	0.150	0.175	0.137	0.211	0.200	0.249
No. Authors	292	82	169	41	67	75	75	75
Observations	7627	3077	6234	1716	2641	2904	2774	2709

Robust standard errors are clustered on the author level and are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
This regression includes quadratic age terms, year dummies, author fixed effects, and a constant term (not reported). There reference period *t* is the year of migration.

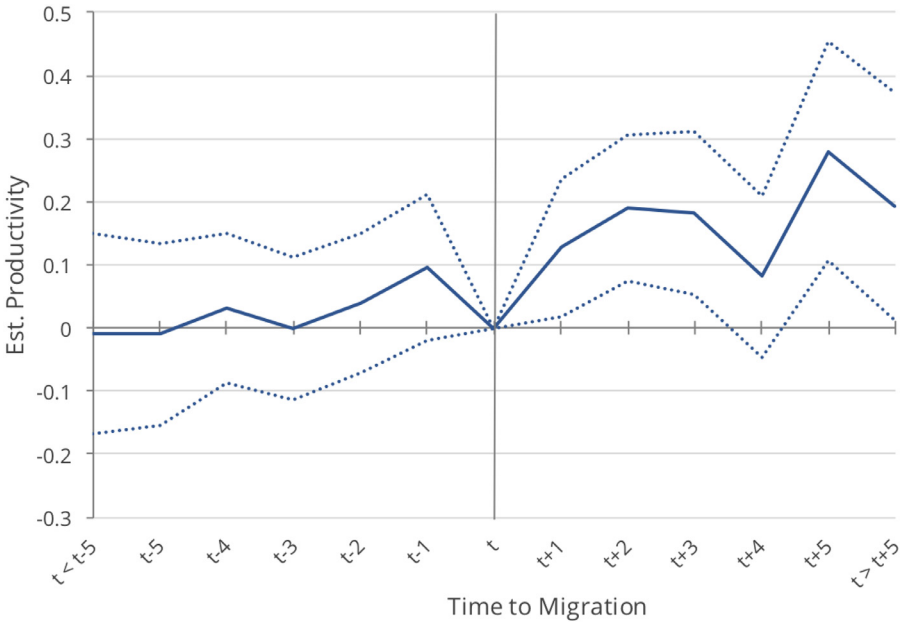

Fig. 8. Estimated trend pre- and post-first migration to London (with 95% CI).

statistically different from productivity in the year of migration. The cohort of authors born in the 18th century receives a statistically significant London premium in the first three years after migrating, while the productivity of the 19th century birth cohort increases in the second and fifth year after migrating. The 20th century birth cohort sees no impact of migrating to London.

In the previous section, there was some evidence of negative selection of the 18th century birth cohort, - the more works an author of this cohort had published, the less likely she was to migrate to London. This finding alone could suggest that the authors in this cohort who were most likely to migrate to London were simply those in the early stages of their careers potentially experiencing an increasing career trajectory leading up to their migration. In this case, the London premium would reflect the continued growth of the careers of successful authors who migrated at an early stage in their career. However, the evidence of this analysis indicates this was not the case. This cohort of authors was

Fig. 9. Estimated trend pre- and post-first migration to London, by century of birth.

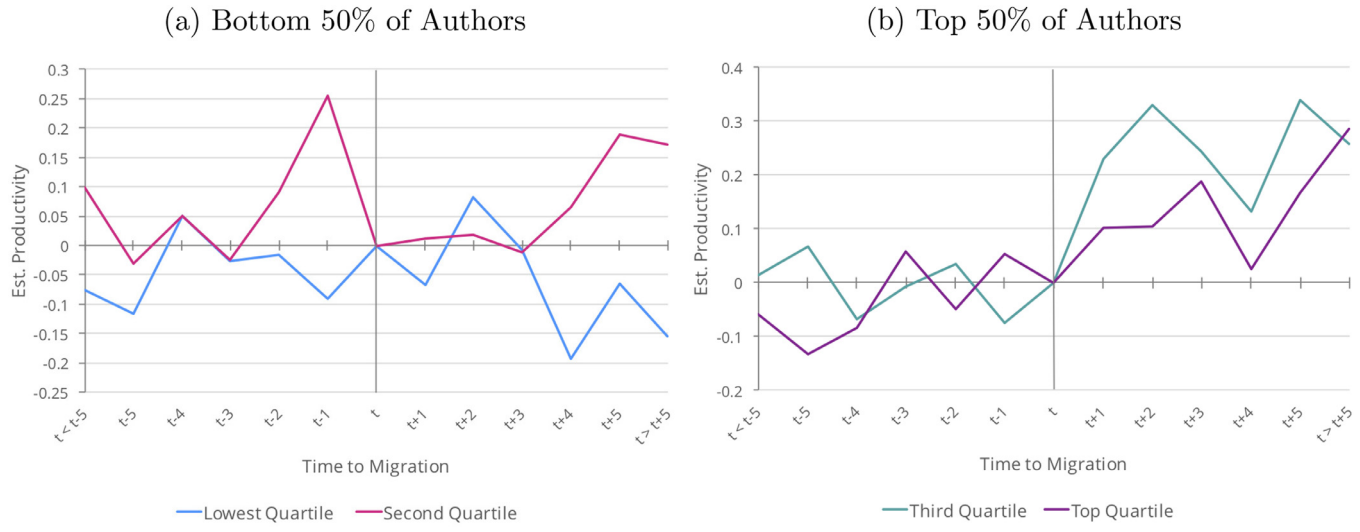

Fig. 10. Estimated trend pre- and post-first migration to London, by impact index quartile.

producing at a relatively constant level of productivity leading to the year of migration, and they experienced a shift in the trajectory of their careers after migrating.

The estimated coefficients for the cohorts of authors in the Impact Index quartiles are illustrated in Fig. 10. The productivity of the cohort of authors ranking in the lowest quartile of the Impact Index is not impacted by migration to London, either before or after the year of migration. The cohort of authors ranking in the second quartile have an increase in productivity in the year before migrating to London relative to the year of migration; however, these authors do not receive a London premium after their arrival.

In contrast, the cohort of authors ranking in the third quartile of the Impact Index sees a statistically significant increase in productivity in the first three years after migrating to London (relative to the year of migration). The productivity of this cohort in the fourth year after migrating to London is not statistically different from the year of migration, but the London premium is observed again from the fifth year. The cohort of authors ranking in the top quartile do not see a statistically significant increase in productivity in the first five years of migration; however, there is some evidence that they may receive gains in the longer term.

Thus, there is no evidence that the London premium is driven by the systematic self-selection of authors who were already experiencing an increasing career trajectory in the years leading up to the migration to London. If this were the case, one would expect to see lower productivity in the years leading up to migration with this trend continuing after migration to London. Instead, productivity appears to be relatively constant in the years leading up to migration with a notable increase in the first few years after migration. The observed decrease in productivity in the fourth year after migration could indicate that migration to London enabled authors to publish works from an existing portfolio instead of (or in addition) to the publishing works that are written after migration before needing to develop a new portfolio of works.

There is no clear evidence to support the possibility that the gains from agglomeration accumulate over time due to learning effects for all cohorts (in contrast to the findings of de la Roca and Puga, 2012). This implies that individual i may not accumulate knowledge or skill

in a way that increases annual productivity in terms of number of works produced. However, it is important to note that the measure of output does not reflect the quality of works. It is possible that learning occurred and that this resulted in higher *quality* works rather than a greater *quantity* of works.

6. Conclusion

This study contributes to current research by empirically analysing the positive externalities associated with the agglomeration of literary activity in historical Britain and Ireland. Specifically, I utilise a unique, purpose-built dataset with information on the birth location and lifetime migration, productivity (in terms of number of publications), and demographic characteristics of 370 authors in the UK and Ireland since 1700. I analyse the migration and clustering of literary artists and found that London was the major literary cluster throughout the sample. I then construct age-productivity profiles to estimate the productivity gains associated with living in a literary cluster.

I find that London emerges as the only major literary cluster, with over 50% of authors locating in London in a given year at its peak concentration. I then estimate the "London effect" utilising individual fixed effects to control for spatial sorting by skill level, and I find that authors who reside in London experience productivity gains of 11.94% per annum compared to their peers living elsewhere – a gain not associated with living in any of the minor literary clusters. I do not find evidence that the results are driven by the systematic self-selection of prolific authors migrating to London or by early-stage authors experiencing a general upward trajectory in their careers that continued after migrating to London. I find heterogeneity of returns to living in London by birth cohort and Impact Index quartile (a measure of author quality), and I find that the cohorts who receive the greatest gains from locating in London are those for which there is the strongest evidence of negative selection with respect to productivity (authors born in the 18th century and authors ranking in the top two quartiles of the Impact Index).

Due to the geographic concentration of creative industries, authors in London likely had access to stronger and more advantageous social networks, in terms of increased connections with their peers (other authors), individuals with influence within the publishing industry (agents, publishers, critics,), and those who are a part of the intellectual and cultural elite (artists, musicians, wealthy patrons). Authors in London also could have taken advantage of the related economic infrastructure and gained from the resulting economies of scale, allowing for a more efficient transformation of ideas into physical book-form.

However, due to the nature of the available data, it is not possible to identify the channels through which authors receive these gains and the degree to which these mechanisms are geographically localised. It is also not possible to determine the kind of agglomeration economy that is present – whether the size of the city (urbanisation) or the concentration of the industry (localisation) plays a larger role in the increase in productivity. Further research and additional data would be required to empirically identify these underlying mechanisms and the type of agglomeration economy.

An important limitation is that the measure of output does not give any indication of the quality of individuals works. This is a critical limitation and deserves consideration, as it is possible that learning occurred and that this resulted in higher *quality* works rather than a greater *quantity* of works. However, capturing quality of output is a challenge that is not unique to this study and is one not easily resolved.

However, the insights into the positive externalities associated with the agglomeration of an industry with few capital requirements are still of relevance to economic researchers, and the contributions to broader understanding of how to access and enhance knowledge spillovers are of interest to both firms and public policy-makers. For example, parallels may be drawn between the creative processes of the historical author and the contemporary software developer. As software development continues to shift from more tradition modes of production to production via remote work, it will be important to understand if (and why) agglomeration of IT firms persists and how individual productivity will be impacted. In this, historical analyses of literary production and other creative industries may be particularly useful.

Acknowledgements

I am grateful to the Grattan Scholar Programme for funding the early stages of this work during my Ph.D. I thank John O'Hagan for his supervision. I thank Christiane Hellmanzik, Lukas Kuld, Ronan Lyons, Vahagn Galstyan, Sanna Nivakoski, Michael O'Grady, Gavin Morrison, Nathaniel Russell, and the participants of the Trinity College Dublin Micro Working Group for their helpful comments on various drafts. I would also like to thank Dr. William Strange and the two anonymous referees for their insight and suggestions. I thank Rahul Dewan for his research assistance.

Appendix A. Detailed data collection methodology

A1. Selection of authors

I aim to identify the centres of co-location (or the geographic clustering of individuals) in order to examine the geography of literary production. In order to do so, I must first systematically determine which individuals to study. In recent years, there have been several attempts to identify significant figures across various fields. Perhaps the most comprehensive of these is the work of Murray (2003), in which Murray identifies and ranks over 4000 of the most globally significant individuals across the arts and sciences from 800 B.C. to 1950. Murray develops his inventories of figures by determining which individuals were consistently included across a large cross-cultural and multilingual collection of sources, including histories, encyclopaedias, and biographical dictionaries. Murray's work includes a roster of 835 significant figures in Western literature. However, Murray's data on individual authors is limited to name, year of birth, national origin, and an index score indicating relative significance. Therefore, additional biographical sources were needed to generate the panel dataset needed for this analysis.

The cultural bias of encyclopaedias, biographical dictionaries, and similar resources has been well documented in works such as Murray (2003), O'Hagan and Kelly (2005), and O'Hagan and Borowiecki (2010). As Murray (2003) explains, this cultural bias is more prominent in literature than other artistic and scientific fields. Additionally, comprehensive literary biographies are quite limited in number. Reference works on literature tend to be specific to one period, movement, and/or language and often include more literary criticism than detailed biographical information. It would require a multi-linguistic effort to systematically and consistently collect biographical information, including information on lifetime location and migration, for all authors. Therefore, due to natural constraints, Murray's inventory of significant figures was not deemed usable for the purposes of this study.

Consequently, I begin by limiting the scope culturally, linguistically, geographically, and temporally to individuals associated with British or Celtic literature born between 1700–1925 with an entry in Encyclopaedia Britannica Online.[29] While there are other countries that are culturally and linguistically similar (e.g. the United States or Canada), transcontinental and transoceanic migration and clustering across such large distances did not occur for most of the sample time period. For the purposes of this study, richness of data over time was preferred to richness of data across geographic reach.

[29] Celtic literature, in this particular usage, is defined as literature associated with Celtic nations within the British Isles (Ireland, Scotland, Isle of Man, Wales, and Cornwall) and does not strictly refer to literature written in the Celtic languages. 'British' literature is not used as an all-encompassing term due to the long and complicated history associated with the concept of 'British identity', the discussion of which lies far beyond the aims of this paper.

I develop a unique inventory of literary artists by collecting every individual associated with British or Celtic literature born between 1700–1925 with an entry in Encyclopaedia Britannica Online. To be defined as a literary artist, an individual must have made at least one unique contribution to poetry or prose, which eliminated individuals whose contributions were strictly limited to translations, textbooks, song-writing, literary criticism, or publishing. Due to posthumous publishing and potential biases that may be associated with an increased interest and analysis in an author immediately following his or her death, living or recently deceased (i.e. post-1999) literary artists were not included. This gave an inventory of 537 literary artists.

Data on the location and migration of the sample of authors was collected from Encyclopaedia Britannica, Literature Online, and The Literary Encyclopaedia. Location information was collected on four levels within the British Isles: City, ceremonial county or lieutenancy area, a modified NUTS 1 regional definition, and country.[30] The modified NUTS 1 regional definition includes all NUTS 1 level statistical regions for England, Scotland, and Wales. England is composed of 9 NUTS 1 regions. Scotland and Wales each compose a single NUTS 1 region. This definition is not used for the island of Ireland, as it lists Northern Ireland and the Republic of Ireland as separate NUTS 1 regions. Historically, these divisions are not so distinctive. Therefore, the traditional provinces of Ireland (Leinster, Ulster, Munster, and Connacht) are used to instead of the NUTS 1 regions. For locations outside the British Isles, information was collected on three levels: city/state, country, and a general category designating distance from British Isles (Rest of Europe, North America, and Rest of World.) In addition, London's city borders have changed dramatically over time, and sources often ambiguously report an individual as "moving to London" rather than specifying that person's exact location. Particularly in the latter part of the sample, the "moving to London" may refer to any part of the metropolitan area. Therefore, I define 'London' as the Greater London ceremonial county rather than the area within the official City of London limits or the London borough.

The roster was reduced further only based on availability of relevant biographical information. To be included in the final roster, authors must have 'near complete' lifetime location information. It was not unusual for individuals to have an unknown location for less than 5 years of their life. Unknown locations were particularly common during periods of conflict for males participating in military service. Such individuals were kept in the sample. With regard to publications, authors either had a comprehensive bibliography listed in Literature Online or a few select works listed across all three online dictionaries. Authors who do not have a comprehensive bibliographical entry were not included in the sample. These exclusions gave a final roster of 370 individuals.

A2. Quality of authors

Murray (2003) uses an advanced column-inch methodology – which relies on the principle that the amount of space devoted to an individual within a source (i.e. total column inches), the more significant that person is – to construct a measure of the relative significance of each person within a field.

Most studies, however, concentrate on developing methods to rank individuals within a specific field. Galenson (2002) uses a textbook illustration method to identify the most significant French artists of the 19th century. This methodology is similar to that of the column-inch but is somewhat more labour-intensive: once the most important sources have been determined, one simply counts the number of illustrations reproduced in each book for the artist. Galenson defines a significant figure as having illustrations in no less than three of his five primary sources and then ranks their significance using the total number of illustrations from a larger sample of 33 art history textbooks. O'Hagan and Kelly (2005) compare the column-inch and textbook-illustration ranking methodologies for 35 visual artists and contrast these results to those of Galenson. They find that both methods are useful for broadly grouping artists by significance and that there is little evidence to suggest that more elaborate ranking methods produce better results.

Verboord (2003) focuses on the relative significance or 'literary prestige' of 502 authors born after 1880. Verboord's measure of significance is composed of a range of indicators within primary areas: number of literary prizes won, entries in literary encyclopaedias (measured by word count rather than column inches), number of academic studies mentioning the author, and the literary reputation of the author's publishing house.[31] Verboord's methodology includes diverse and comprehensive indicators, but it is not possible to use all of these indicators to estimate literary prestige for earlier time periods. Consistent and comprehensive information on authors' publishing houses is difficult to obtain, particularly information from before the mid-1800s. Furthermore, it is not possible to use the number of literary prizes won for earlier time periods, as literary prizes are quite a recent invention. Literary prizes were limited to schools and universities until the mid-nineteenth century, and the earliest national literary prize in Great Britain was not founded until 1919 (Shaffer, 2008).

It should be noted that while those in the literary and wider art world are often hesitant to attach ordinal rankings to artists or to designate one author as indisputably "better than" another, the previously discussed methodologies do not create an arbitrary cardinal measure of quality (e.g. the stars rating system generally associated with movie ratings.) Instead, such rankings simply reflect the amount of critical attention given to one author relative to her peers. Furthermore, previous studies have shown that systematic rankings based on measurable attributes can provide comparable judgment to experts in the field and, in some cases, may even outperform them. Throsby (1994), for example, found that the aesthetic judgments of art experts are not random but systematically based on observable characteristics such as artist's school of work, career stage, and past achievement.

The methodology used to determine significance in this study is similar to that of Verboord (2003); though, the indicators in this study are limited to entries in literary encyclopaedias and attention given in literary studies. Similar to the measures of significance used in other studies, the Impact Index simply reflects the amount of contemporary critical attention an individual author has received relative to her peers.

Sources of information on the biographies and critical reception of the authors was limited to three online encyclopaedias: Encyclopaedia Britannica (2014), The Literary Encyclopaedia (2014), and Literature Online (2014). In addition to providing a biographical entry for each author, Literature Online also contains a database of literary criticism and reference works, such as ABELL (Annual Bibliography of English Language and Literature 1920-), MLA International Bibliography, an electronic library of 392 full-text journals of literary criticism, an indexed directory of 4400 periodicals, and a number of other reference works. Each author's main page includes links to journal articles, citations, and reference articles about that author. Journals of literary criticism include peer-reviewed academic articles on new research, reviews or critiques of previous works, and book reviews. The criticism component of the Impact Index was created by recording the total number of entries under each author's respective criticism page.

For the word-count component of the Impact Index, total word count rather than column-inch method is used because differences in font and format made cross-resource column-inch comparisons infeasible. Additionally, the absolute word count is used rather than a measure weighted by the total length of the source. The length of an entry relative to the to-

[30] NUTS 1 statistical regions are the *Nomenclature of Territorial Units for Statistics, Level 1* regional standards developed by the European Union.

[31] The literary reputation of the author's publishing house is determined by the total number of prestigious authors on the publisher's list. This definition assumes that the prestige of the publishing house is likely to influence the prestige of the individual author.

tal length of each encyclopaedia could not be determined because there exists no simple method of determining the complete length of these online resources due to the format of the online systems. Furthermore, due to the nature of online hosting, publishers do not experience the same space limitations associated with physical encyclopaedias, and so the authors of each biographical entry has greater freedom to expand.

References

Audretsch, B., 1998. Agglomeration and the location of innovative activity. Oxford Rev. Econ. Policy 14 (2), 18–29.

Audretsch, D.B., Feldman, M.P., 1996. R&D spillovers and the geography of innovation and production. Am. Econ. Rev. 630–640.

Audretsch, D.B., Feldman, M.P., 2004. Knowledge spillovers and the geography of innovation. In: Cities and Geography. Vol. 4 of Handbook of Regional and Urban Economics. Elsevier, pp. 2713–2739. doi:10.1016/S1574-0080(04)80018-X.

Azoulay, P., Zivin, J.S., Wang, J., 2010. Superstar extinction. Q. J. Econ. 125 (2), 549–589.

Borowiecki, K.J., 2012. Are composers different? historical evidence on conflict-induced migration (1816–1997). Eur. Rev. Econ. Hist. 16 (3), 270–291.

Borowiecki, K.J., 2013. Geographic clustering and productivity: an instrumental variable approach for classical composers. J. Urban Econ. 73 (1), 94–110.

Borowiecki, K.J., O'Hagan, J.W., 2013. Impact of war on individual life-cycle creativity: tentative evidence in relation to composers. J. Cultural Econ. 37 (3), 347–358.

Boschma, R., 2005. Proximity and innovation: a critical assessment. Reg. Stud. 39 (1), 61–74.

Carlino, G., Kerr, W.R., 2015. Agglomeration and innovation. In: Gilles Duranton, J.V.H., Strange, W.C. (Eds.), Handbook of Regional and Urban Economics. Vol. 5 of Handbook of Regional and Urban Economics. Elsevier, pp. 349–404. doi:10.1016/B978-0-444-59517-1.00006-4.

Carlino, G.A., Chatterjee, S., Hunt, R.M., 2007. Urban density and the rate of invention. J. Urban Econ. 61 (3), 389–419.

Ciccone, A., Hall, R.E., 1996. Productivity and the density of economic activity. Am. Econ. Rev. 86 (1), 54.

Combes, P.-P., Duranton, G., Gobillon, L., 2008. Spatial wage disparities: sorting matters!. J. Urban Econ. 63 (2), 723–742.

Combes, P.-P., Gobillon, L., 2015. The empirics of agglomeration economies. In: Gilles Duranton, J.V.H., Strange, W.C. (Eds.), Handbook of Regional and Urban Economics. Vol. 5 of Handbook of Regional and Urban Economics. Elsevier, pp. 247–348. doi:10.1016/B978-0-444-59517-1.00005-2.

D'Este, P., Guy, F., Iammarino, S., 2013. Shaping the formation of university – industry research collaborations: what type of proximity does really matter? J. Econ. Geogr. 13 (4), 537–558.

Duranton, G., Puga, D., 2004. Micro-foundations of urban agglomeration economies. In: Cities and Geography. Vol. 4 of Handbook of Regional and Urban Economics. Elsevier, pp. 2063–2117. doi:10.1016/S1574-0080(04)80005-1.

Encyclopaedia Britannica, 2014. www.britannica.com.

Erickson, L., 2002. "Unboastful bard": originally anonymous english romantic poetry book publication, 1770–1835. New Lit. Hist. 33 (2), 247–278.

Farrell, M.P., 2003. Collaborative Circles: Friendship Dynamics and Creative Work. University of Chicago Press.

Galenson, D.W., 2002. Quantifying artistic success: ranking french paintersand paintings-from impressionism to cubism. Hist. Methods 35 (1), 5–19.

Gertler, M.S., 2003. Tacit knowledge and the economic geography of context, or the undefinable tacitness of being (there). J. Econ. Geogr. 3 (1), 75–99.

Giorcelli, M., Moser, P., 2015. Copyright and creativity: evidence from Italian operas. Working Paper.

Glaeser, E., Maré, D., 2001. Cities and skills. J. Labor Econ. 19 (2). 316–42

Glaeser, E.L., Kallal, H.D., Scheinkman, J.A., Shleifer, A., 1992. Growth in cities. J. Polit. Econ. 100 (6). 1126–52

Great Britain Historical GIS/University of Portsmouth, 2017. A vision of Britain through time: Population statistics, total population. http://www.visionofbritain.org.uk/unit/10001043/cube/TOT_POP.

Hellmanzik, C., 2010. Location matters: estimating cluster premiums for prominent modern artists. Eur. Econ. Rev. 54 (2), 199–218.

Hellmanzik, C., 2013. Democracy and economic outcomes: evidence from the superstars of modern art. Eur. J. Polit. Econ. 30, 58–69.

Helsley, R.W., Strange, W.C., 2002. Innovation and input sharing. J. Urban Econ. 51 (1), 25–45.

Howells, J., 2012. The geography of knowledge: never so close but never so far apart. J. Econ. Geogr. 12 (5), 1003–1020.

Jaffe, A.B., Trajtenberg, M., Henderson, R., 1993. Geographic localization of knowledge spillovers as evidenced by patent citations. Q. J. Econ. 108 (3), 577–598.

Krugman, P., 1991. Increasing returns and economic geography. J. Polit. Econ. 99 (3), 483–499.

Levin, S.G., Stephan, P.E., 1991. Research productivity over the life cycle: evidence for academic scientists. Am. Econ. Rev. 114–132.

Literature Online, 2014. www.lion.chadwyck.co.uk.

Marshall, A., 1890. Principles of Economics. Macmillan, London.

Mokyr, J., 2005. Mobility, creativity, and technological development: david hume, immanuel kant and the economic development of europe. Northwestern University, unpublished manuscipt.

Moser, P., Voena, A., Waldinger, F., 2014. German jewish émigrés and us invention. Am. Econ. Rev. 104 (10), 3222–3255.

Murray, C.A., 2003. Human Accomplishment: The Pursuit of Excellence in the Arts and Sciences, 800 BCE to 1950. HarperCollins New York.

National Archives of Ireland, 2016. About pre-1901 census fragments. http://www.census.nationalarchives.ie/help/pre1901.html.

Office for National Statistics, 2015. UK population estimates 1851 to 2014 (ad hoc table no. 004356). https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/adhocs/004356ukpopulationestimates1851to2014.

Office for National Statistics London Datastore, 2014. Historical Census Population. https://data.london.gov.uk/dataset/historic-census-population?resource=2c7867e5-3682-4fdd-8b9d-c63e289b92a6.

O'Hagan, J., Borowiecki, K.J., 2010. Birth location, migration, and clustering of important composers: historical patterns. Hist. Methods 43 (2), 81–90.

O'Hagan, J., Kelly, E., 2005. Identifying the most important artists in a historical context: methods used and initial results. Hist. Methods 38 (3), 118–125.

Online Historical Population Reports, 2016. Census of Great Britain, 1851, Population tables, I. Number of the inhabitants in 1801, 1811, 1821, 1831, 1841 and 1851. Vol. I BPP 1852-53 LXXXV (1631), pages 68–69. hhtp://www.histpop.org.

Rallet, A., Torre, A., 1998. On geography and technology: The case of proximity relations in localized innovation networks. In: Steiner, M. (Ed.), Clusters and Regional Specialisation. Pion Publication.

de la Roca, J., Puga, D., 2012. Learning by working in big cities. CEPR Discussion Papers, 9243. C.E.P.R. Discussion Papers.

Rose, M., 1988. The author as proprietor: donaldson v. becket and the genealogy of modern authorship. Representations (23) 51–85.

Shaffer, B.W., 2008. A Companion to the British and Irish novel 1945–2000, 45. John Wiley & Sons.

St Clair, W., 2004. The Reading Nation in the Romantic Period. Cambridge University Press.

Ter Wal, A.L., Boschma, R.A., 2009. Applying social network analysis in economic geography: framing some key analytic issues. Ann. Reg. Sci. 43 (3), 739–756.

The Literary Encyclopaedia, 2014. http://www.litencyc.com.

Throsby, D., 1994. The production and consumption of the arts: a view of cultural economics. J. Econ. Lit. 32 (1), 1–29.

Torre, A., Rallet, A., 2005. Proximity and localization. Reg. Stud. 39 (1), 47–59.

Upchurch, A., 2004. John maynard keynes, the bloomsbury group and the origins of the arts council movement. Int. J. Cultural Policy 10 (2), 203–217.

US Department of Justice and the Federal Trade Commission, 2010. Horizontal merger guidelines. https://www.justice.gov/atr/horizontal-merger-guidelines-08192010.

Verboord, M., 2003. Classification of authors by literary prestige. Poetics 31 (3), 259–281.

Von Hippel, E., 1994. Sticky information and the locus of problem solving: implications for innovation. Manag. Sci. 40 (4), 429–439.

Waldinger, F., 2010. Quality matters: the expulsion of professors and the consequences for Ph.D student outcomes in nazi germany. J. Polit. Econ. 118 (4), 787–831.

Waldinger, F., 2012. Bombs, Brains, and Science: The Role of Human and Physical Capital for the Creation of Scientific Knowledge. Technical Report. Competitive Advantage in the Global Economy (CAGE).

Waldinger, F., 2012. Peer effects in science: evidence from the dismissal of scientists in nazi germany. Rev. Econ. Stud. 79 (2), 838–861.

Wooldridge, J.M., 2010. Count, fractional, and other nonnegative responses. Econometric analysis of cross section and panel data, 2 MIT Press.