London Calling? Agglomeration Economies in Literature since 1700

Sara Mitchell [*]

Trinity College Dublin

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. We find the average author experiences productivity gains of 10.68% per annum when residing in London, controlling for spatial sorting by skill level. While we do not find evidence of dynamic learning effects, we confirm that our results are not driven by a dynamic spatial self-selection process or authors who strategically build a portfolio of works before moving to London and publishing in quick succession.

Keywords: Geographic concentration, cities, mobility, productivity, urban history, literary artists

JEL Classifications: D24, J24, J61, N90, R11, Z11

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 dynamics for industries

[*]I am grateful to the Grattan Scholar Programme for funding my PhD research. I thank John O'Hagan for his supervision. I thank Ronan Lyons, Vahagn Galstyan, Sanna Nivakoski, Michael O'Grady, Gavin Morrison and the participants of the TCD Micro Working Group for their helpful comments on various drafts. I thank Rahul Dewan and Rónán O'Connor for their research assistance.

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. Authors are also 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? We argue that tacit knowledge exchange drives the geographic clustering of authors and that these flows cause a localised increase in author productivity. In order to answer these questions, we determine the cities in which literary activity clusters, and we empirically estimate these returns to literary agglomeration.

Specifically, we 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. We begin by discussing the data collection methodology. We then analyse the patterns of migration and clustering of authors, and we construct age-productivity profiles to determine the productivity gains (if any) associated with the geographic clustering of literary activity. Following Davis and Dingel (2012), we acknowlege productivity gains from agglomeration effects (i.e. via tacit knowlege transfer) operate as both a by-product and driver of geographic clustering of literary production. We utilise individual fixed effects to control for spatial sorting by skill level, and the within estimator reveals the average author annual productivity is 10.68% per annum higher when the author resides in London, the only major literary cluster during the time period. While we do not find evidence of dynamic learning effects, we find no evidence that our results are driven by a dynamic self-selection process. We also find no evidence that authors strategically build a portfolio of works before moving to London and publishing them in quick succession.

The remainder of this paper is structured as follows. In the next section, we outline the theoretical framework by discussing the relationship between geographic proximity and tacit knowledge, agglomeration economies and recent empirical evidence. Next, we discuss the data collection methodology and descriptive statistics. Section 4 outlines the identification strategy, and results are presented in Section 5. Section 6 concludes.

2 Theoretical Framework

2.1 Geographic proximity and tacit knowledge

One of the earliest economic explanations of geographic clustering of industry was posited by Alfred Marshall:

> *"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, inventions and improvements in machinery, in processes and the general organization of the business have their merits promptly discussed: 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. And presently subsidiary trades grow up in the neighbourhood, supplying it with implements and materials, organizing its traffic, and in many ways conducing to the economy of its material."* (Marshall 1890, p. 225)

More formally, 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).

Duranton and Puga (2004) discuss the micro-foundations of Marshallian agglomeration economies and divide them into broad types based on the mechanisms that drive their formation: sharing, matching, and learning. 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 has been 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.[1] 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). [2]

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').[3] 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.

Authors are anecdotally described as solitary and independent workers, so authorship could be an inherently individual process and thus authors would not benefit to the same degree (if at all) from agglomeration. However, authors are cognitively proximate by the nature of their profession, and it is likely a sample of authors will be socially or institutionally proximate if the sample is limited to a relatively small geographic region, as they are likely to share a common language and culture and interact with the same members of a wider literary-industrial complex (publishers, editors, critics, etc.) Thus, tacit knowledge transfer is possible in this context. Given these necessary conditions for tacit knowledge transfer, we hypothesise that productivity gains from geographic co-location can (and do) exist for literary production.

[1]See Audretsch, 1998; Gertler, 2003; Von Hippel, 1994 .

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

[3]Also see Mokyr, 2005; D'Este et al., 2013; Torre and Rallet, 2005; Boschma, 2005; Hellmanzik, 2013; Audretsch, 1998; Rallet and Torre, 1998.

2.2 Agglomeration economies and productivity

Beyond driving the formation of agglomeration economies, location and proximity also play important roles in increasing productivity. Geographic proximity increases the potential for tacit knowledge transfer which then, in turn, increases firm or individual productivity (Combes and Gobillon, 2015). These gains attract new firms that who relocate to exploit the gains, further increasing the geographic concentration of innovative activity. Chatterji et al. (2013) find decreasing returns to such spillovers, while Ter Wal and Boschma (2009) warn of cognitive lock-in – the risk of getting 'locked in' to established ways of thinking instead of continuing to develop new ideas increases as the density of a cluster increases (also see Boschma, 2005; Jones, 2009; Audretsch, 1998). This cyclical relationship between location, proximity, and knowledge spillovers also makes identification of causal relationships difficult (Audretsch and Feldman, 2004). Porter (2011) discusses how competition and rivalry due to geographic proximity increase overall standards and the probability competing firms or individuals will be considered 'prominent.' Kelly and O'Hagan (2007), however, warns that the direction of the clustering/prominence relationship is not clear and may even be cyclical: some prominent agents may choose to cluster to reap the benefits of doing so, while others become prominent because of the gains they received from working in a cluster.

Furthermore, there have also been attempts to estimate the gains from economic clustering across the creative industries and understand how these dynamics change by field. Perhaps the most prevalent method of studying the gains from geographic proximity is by "following the paper trail" in analysing the spatial clustering of patent registrations from R&D by private firms and universities. (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.) 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."

A more recent set of literature analyses agglomeration and knowledge spillovers by studying the individual scientist, often in a historical context. Waldinger (2012a) evaluates the productivity losses (in terms of academic publications rather than patents) associated with the breakdown of research clusters (i.e. university science departments), utilising the dismissal of mostly Jewish scientists and the destruction of facilities by Allied bombing in Nazi Germany as exogenous shocks. He finds significant decreases in departmental productivity, due to both human capital and physical capital shocks. Waldinger (2012b, 2010) conduct similar studies on university mathematics and science departments in Nazi Germany, finding 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 PhD students within their department.

Another new body of literature looks at the clustering of broadly-defined creative activity. Borowiecki (2013) conducts a historical study examining the impact of geographic clustering on the productivity of classical composers. He finds that clustering has a positive and significant impact on composer productivity and finds heterogeneity of returns across clusters and composer quality. Borowiecki (2012) and Borowiecki and O'Hagan (2013) follow this by analysing the impact of war on clustering and life-cycle creativity. They find that the share of composers drops during periods of conflict, composer productivity is negatively impacted by the presence of a conflict, and the outward migration of composers significantly reduces a country's long-term creative potential. The gains from the clustering of prominent visual artists is analysed by Hellmanzik (2010), who finds that the artwork of prominent visual artists is valued higher when produced in a creative cluster. She also finds heterogeneity of returns by cluster location and artist quality.

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.[4] 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

[4]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.

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). This 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. We rely on observable measures of productivity (output per annum) for our analysis. Output is measured in terms of number of known publications rather than number of works ever written.[5] Of the roster of 537 significant figures, only 370 have complete or near-complete lifetime location and publication information.[6] This sample of 370 authors is used in the econometric analysis in the following sections.

Following the methodology of Murray (2003) and Verboord (2003), we 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. 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 30 literary artists and their respective Impact Index value is listed in Table 3 in Appendix B.2.

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

3.2 Migration Trends and Summary Statistics

The Greater London Area is the location with the most births of literary artists, as 79 (or 21.35%) of the artists were born within this region. (See Table 1 in Appendix B.2.) We also

[5] Publications include any type of publication: novel, collection of poetry or short stories, anthologies, contributions to literary magazines, plays, memoirs, etc. All publications were given equal weight. We have no measure of the quality of individual publications.

[6] 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.

evaluate spatio-temporal trends in lifetime movement to determine if authors tend to migrate and cluster in certain locations. As seen in Figure 1, clustering intensity is quite high throughout the sample. London[7] consistently emerges as the largest literary cluster, with over 30% of all authors in a given year residing in London until the end of the 19th century.

At its peak, London was home to over 50% of all authors. Dublin, Edinburgh, Oxford, and Cambridge emerge as the only other cities that 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 clustering in each city per annum), we define London as the only literary cluster for analysis in this study. 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. 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.

We also investigate the variation in movement in and out of London, as a key concern is that London arises as a cluster 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 clustering. This is significant in two ways. First, 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. Second, while the important of London as a literary cluster is not surprising, the fact that centres of learning and culture emerge as the only other city-level clusters rather than simply large, urban area is revealing. Gains from agglomeration of literary activity will not be due to economies of scale effects associated with

[7] London is defined as the Greater London ceremonial county rather than the area within the official City of London limits, as 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 "London area" may refer to any part of the metropolitan area.

general production in a large city. Instead, it suggests that having a diverse intellectual elite may be a more important factor driving migration and clustering, supporting the tacit-knowledge exchange hypothesis.

Summary statistics for the sample of 370 authors are presented in Table 2. Authors tended to be engaged in work-related activities for a great portion of their life, on average publishing their first work at age 25 and continuing to work for just under 40 of their 66 years of life. Over the course of their careers, authors produced an average of 31 publications, publishing an average of 0.88 works per year. Over their lifetimes, authors spent around 21 years in London. There are some distinct gender differences, particularly with regard to total lifetime works with male authors publishing around 10 more works over the course of their lifetimes.

Figure 2 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 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.[8] The dynamics of publishing are further detailed in Figure 3, 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

We 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:

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

[8]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).

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 Equation 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, Equation 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. [9] 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 individual fixed-effects can adequately address the issue of workers sorting across locations on unobservables. [10]

We 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. We 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.[11] 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

[9] For example, 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.

[10] 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."

[11] We 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. We have no measure of the quality of individual publications.

in the same way. We also include a full set of author fixed effects, given by σ_i.

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 Figure 4. Therefore, we limit our study to potential active years, which we define as age 16 and older. This represents the lower bound for age at first publication. Additionally, many authors in our 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 our analysis begins in 1725 (the first year with multiple observations) and ends in 1999 (when the last authors died). We also supplement our analyses with a negative binomial model, which is a more appropriate specification for over-dispersed count data.[12]

5 Results

The estimates for Equation 2 are presented in Table 4. 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), we 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. Column (3) includes the coefficients for the negative binomial model with individual fixed effects, and Column (4) provides the respective incident rate ratios. In all four 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 explain by location. However, when individual fixed effects are introduced in Column (2), an author residing in London experience 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 0.88, this translates to a 10.68 % increase in annual productivity. The negative binomial incident rate ratios (NB IRR) reported in Column (4) indicate that an author sees a 24.3% increase in

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

the probability of publishing in a given year while residing in London compared to years of her life when residing elsewhere. In all specifications, location is not correlated with productivity in the cross-section, but individual authors do experience productivity gains from locating in a geographic cluster throughout their life-cycles. If there existed a correlation between location and productivity in the cross-section, this would suggest that standard large-city economy of scale effects are a primary underlying factor. The fact that there is only a significant relationship within author observations suggests that there exists some sort of individual learning process associated with the geographic clustering of literary activity, which supports the tacit-knowledge hypothesis presented previously.

In all specifications, age has a positive but diminishing effect, which is consistent with general findings in the literature on individual productivity. (See, for example, Levin and Stephan (1991).) 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).)

5.1 Robustness Analysis

5.1.1 Spatial Selection as as 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 4 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'), the fixed effects estimates will be a lower bound. We explicitly address this issue by using a logit model to determine the role of past success (in terms of output) in the probability of migrating to London.[13]

In Table 5, we present logit regressions predicting the probability of migrating to London to examine the role of past success in determining migration. In Column (1), we use the previous year's output to examine if immediate past success increases the probability of migration. In Column (2), we use the total output to date to determine if cumulative success increases the probability of migration. In both specifications, previous success has no statistically significant impact on the probability of migrating to London. The coefficients on both output measures are negative, suggesting that greater past success decreases the probability of migrating. We repeat these specifications by quality level (low/high[14]) and by century of birth to determine if these results are being driven by a particular cohort. In both specifications, the results are consistent (see Tables 6 and 7 in the Appendix, respectively). Thus, we conclude that our results on the gains from the geographic clustering of literary activity in London are not driven by the systematic self-selection of recently successful authors.

5.1.2 Duration Effects

In our previous analysis, the regression specifications have the explicit assumption that individuals receive an instantaneous productivity gain upon arrival in London with no change to the rate of productivity. However, De la Roca and Puga (2012) argue that this approach to agglomeration is too simplistic. Individuals receive a static, instantaneous gain from agglomeration (e.g. from immediate access to better local resources), but they may also receive a dynamic gain reflecting learning that accumulates over time. De la Roca and Puga (2012) find that the magnitude of this dynamic gain is positively related to city size, and they argue that the presence of dynamic agglomeration effects is clear evidence of learning and tacit knowledge spillovers.

[13]We use 25-year dummies rather than year dummies for the logit models to be parsimonious with the degrees of freedom.

[14]See Section 5.4 for a detailed discussion of the quality variable.

This theoretical distinction is illustrated in Figure 5. The solid black (lowest) line represents the age-productivity profile of an individual who does not live in London. The solid red (middle) line represents the age-productivity profile of an individual upon moving to London under the assumption that there is only a static gain. The dashed red (upper) line represents both the fixed gain and the dynamic gain. Mathematically, the static gain is represented by an increase in the intercept of the original age-productivity function, and the dynamic gain is represented by an increase in the slope of the original age-productivity function.

This graphical analysis is also useful in assessing another potential issue. It is possible that authors may build a large portfolio of writings before moving to London. Upon arrival, the author may make connections with local literary agents and publishers who then publish a number of the author's works in quick succession. In this case, the estimate for *London* will be an overestimate of the true knowledge spillover effects as it would also capture the fact that authors strategically withhold writings before 'dumping' them on the market within the first years of arriving in London. Although the average London effect indicates a increase in productivity relative to the periods the author does not reside in London, the author would actually be receiving a sharp spike in productivity immediately after arrival in London before experiencing a drop to at or near pre-migration levels.

Following the methodology of De la Roca and Puga (2012), we consider the following extension in order to simultaneously estimate the dynamic and static agglomeration effect of living in London:

$$output_{it} = \delta London_{it} + \gamma_1 Exper_{it} + \gamma_2 Exper_{it}^2 + \beta_1 age_{it} + \beta_2 age_{it}^2 + \sigma_i + \mu_t + \epsilon_{it} \qquad (3)$$

in which $Exper_{it}$ is defined as the number of consecutive years individual i has spent in London since the year of migration to London, and $Exper_{it}^2$ allows for diminishing returns to duration in London. The results are provided in Table 8. The net duration effect, which is found by estimating this model without *London*, is given in Column (1). The breakdown of this effect into static and dynamic effects is shown in Column (2): the estimate for *London* gives the fixed effect and the estimate for *Experience* gives the dynamic effect. Life-cycle aging effects added in Column (3). The corresponding negative binomial estimates for these models are given in

Columns (4), (5) and (6), respectively.

Plotting the point estimates allows us to visually detect the existence of 'dumping' while also determining if gains from agglomeration accumulate. We re-estimate the full model specified by Equation 3 experience dummies rather than an experience trend using OLS with author fixed effects.[15] Figure 6 provides an illustration of this post-migration premium by static, dynamic and net agglomeration effects. In this graph, the London premium in time 10 is interpreted as the productivity of individual i in the 10th consecutive year in London after migrating relative to the years she was not in London.

As can be seen in Table 8 and in Figure 6 (and in contrast to the findings of De la Roca and Puga (2012)), there is little evidence to support the hypothesis that the gains from agglomeration accumulate over time above life-cycle aging effects. This implies that individual i does not accumulate knowledge or skill in a way that increases annual productivity in terms of *number of works produced*. However, it is possible that the author accumulates skills that instead result in producing *higher quality work*. Data on the quality of work is unavailable, so we are not able to address this issue in further detail in this context. There is also no evidence of strategic 'dumping' within the first few years of arrival in London. Authors receive relatively constant gains for the duration of their stay in London.

This exercise is also useful for visualising the net agglomeration gains by comparing the pre- and post-migration trends. To do this, we re-estimate Equation 3 with dummies for duration in London and for each of the 5 years before migrating to London. These estimates are illustrated in Figure 7. There is a distinct jump in productivity once the individual moves to London. Relative to the periods in which individual i is not in London, the gains from living in London are essentially the same whether she has lived in London for 2 consecutive years of 10 consecutive years. However, there are no net gains in the year of migration. It appears individual i does undergo an initial learning process when she first moves to London before the gains are fully realised, but then the individual does not accumulate further gains over time.

[15]These results are not reported but are consistent with those in Table 8.

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, we 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. We analyse the migration and clustering of literary artists and found that London was the major literary cluster throughout the sample. We then construct age-productivity profiles to estimate the productivity gains associated with the geographic clustering of literary activity.

We find that authors who reside in London experience productivity gains of 10.68% per annum compared to their peers living elsewhere. These results are not driven by self-selection of recently successful authors in migrating to London. This increase in productivity is primarily due to a static agglomeration effect and not due to a dynamic learning effect. There is some evidence of a short adjustment period when individuals in the first year of arrival in London, but productivity gains do not accumulate thereafter. There is also no evidence that authors strategically withhold writings until migrating to London and then 'dump' them on the market.

The geographic concentration of creative workers is certain to have played an important role in the transfer and diffusion of knowledge, as well as the generation of new knowledge and ideas. 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, we conclude that further research is needed to fully understand channels through which authors receive these gains and the degree to which these mechanisms are geographically localised.

Due to the nature of the data used in this study, broad generalisations of these results are limited. 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.

References

Audretsch, B. (1998). Agglomeration and the location of innovative activity. *Oxford Review of Economic Policy*, 14(2):18–29.

Audretsch, D. B. and Feldman, M. P. (1996). R&D spillovers and the geography of innovation and production. *American Economic Review*, pages 630–640.

Audretsch, D. B. and Feldman, M. P. (2004). Knowledge spillovers and the geography of innovation. In *Cities and Geography*, volume 4 of *Handbook of Regional and Urban Economics*, pages 2713 – 2739. Elsevier.

Azoulay, P., Zivin, J. S., and Wang, J. (2010). Superstar extinction. *Quarterly Journal of Economics*, 125(2):549–589.

Borowiecki, K. J. (2012). Are composers different? Historical evidence on conflict-induced migration (1816–1997). *European Review of Economic History*, 16(3):270–291.

Borowiecki, K. J. (2013). Geographic clustering and productivity: An instrumental variable approach for classical composers. *Journal of Urban Economics*, 73(1):94–110.

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

Boschma, R. (2005). Proximity and innovation: a critical assessment. *Regional Studies*, 39(1):61–74.

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

Chatterji, A., Glaeser, E. L., and Kerr, W. R. (2013). Clusters of entrepreneurship and innovation. Technical report, National Bureau of Economic Research.

Ciccone, A. and Hall, R. E. (1996). Productivity and the density of economic activity. *American Economic Review*, 86(1):54.

Combes, P.-P., Duranton, G., and Gobillon, L. (2008). Spatial wage disparities: Sorting matters! *Journal of Urban Economics*, 63(2):723–742.

Combes, P.-P. and Gobillon, L. (2015). The empirics of agglomeration economies. In Gilles Duranton, J. V. H. and Strange, W. C., editors, *Handbook of Regional and Urban Economics*, volume 5 of *Handbook of Regional and Urban Economics*, pages 247 – 348. Elsevier.

Davis, D. R. and Dingel, J. I. (2012). A spatial knowledge economy. Technical report, National Bureau of Economic Research.

De la Roca, J. and Puga, D. (2012). Learning by working in big cities.

D'Este, P., Guy, F., and Iammarino, S. (2013). Shaping the formation of university – industry research collaborations: What type of proximity does really matter? *Journal of Economic Geography*, 13(4):537–558.

Duranton, G. and Puga, D. (2004). Micro-foundations of urban agglomeration economies. In *Cities and Geography*, volume 4 of *Handbook of Regional and Urban Economics*, pages 2063 – 2117. Elsevier.

Encyclopaedia Britannica (2014). `www.britannica.com`.

Galenson, D. W. (2002). Quantifying artistic success: Ranking French paintersand paintingsfrom Impressionism to Cubism. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 35(1):5–19.

Gertler, M. S. (2003). Tacit knowledge and the economic geography of context, or the undefinable tacitness of being (there). *Journal of Economic Geography*, 3(1):75–99.

Glaeser, E. and Maré, D. (2001). Cities and skills. *Journal of Labor Economics*, 19(2):316–42.

Glaeser, E. L., Kallal, H. D., Scheinkman, J. A., and Shleifer, A. (1992). Growth in cities. *Journal of Political Economy*, 100(6):1126–52.

Hellmanzik, C. (2010). Location matters: Estimating cluster premiums for prominent modern artists. *European Economic Review*, 54(2):199–218.

Hellmanzik, C. (2013). Democracy and economic outcomes: Evidence from the superstars of modern art. *European Journal of Political Economy*, 30:58–69.

Howells, J. (2012). The geography of knowledge: Never so close but never so far apart. *Journal of Economic Geography*, 12(5):1003–1020.

Jaffe, A. B., Trajtenberg, M., and Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *Quarterly Journal of Economics*, 108(3):577–598.

Jones, B. F. (2009). The burden of knowledge and the death of the renaissance man: Is innovation getting harder? *Review of Economic Studies*, 76(1):283–317.

Kelly, E. and O'Hagan, J. (2007). Geographic clustering of economic activity: The case of prominent Western visual artists. *Journal of Cultural Economics*, 31(2):109–128.

Krugman, P. (1991). Increasing returns and economic geography. *Journal of Political Economy*, 99(3):483–499.

Levin, S. G. and Stephan, P. E. (1991). Research productivity over the life cycle: Evidence for academic scientists. *American Economic Review*, pages 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., and Waldinger, F. (2014). German jewish émigrés and us invention. *American Economic Review*, 104(10):3222–3255.

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

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

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

Porter, M. E. (2011). *Competitive advantage of nations: creating and sustaining superior performance.* Simon and Schuster.

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

Shaffer, B. W. (2008). *A companion to the British and Irish novel 1945-2000*, volume 45. John Wiley & Sons.

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

Ter Wal, A. L. and Boschma, R. A. (2009). Applying social network analysis in economic geography: framing some key analytic issues. *The Annals of Regional Science*, 43(3):739–756.

The Literary Encyclopaedia (2014). `www.litencyc.com`.

Throsby, D. (1994). The production and consumption of the arts: A view of cultural economics. *Journal of Economic Literature*, 32(1):1–29.

Torre, A. and Rallet, A. (2005). Proximity and localization. *Regional Studies*, 39(1):47–59.

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. *Management Science*, 40(4):429–439.

Waldinger, F. (2010). Quality matters: The expulsion of professors and the consequences for PhD student outcomes in Nazi Germany. *Journal of Political Economy*, 118(4):787–831.

Waldinger, F. (2012a). 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. (2012b). Peer effects in science: Evidence from the dismissal of scientists in Nazi Germany. *Review of Economic Studies*, 79(2):838–861.

A Detailed Data Collection Methodology

A.1 Selection of Authors

We 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, we 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 4,000 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. Howver, 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 need 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, we 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.[16] 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.

We 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

[16] 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.

contribution to poetry or prose, which eliminated individuals whose contributions were strictly limited to translations, textbooks, songwriting, 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.[17] 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.)

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.

A.2 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

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

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.[18] 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-ninteenth 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 judgement to experts in the field and, in some cases, may even outperform them. Throsby (1994), for example, found that the aesthetic judgements 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

[18]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.

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 4,400 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 total length of each encyclodaedia 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.

B Tables and Figures

B.1 Figures

Figure 1: Share of Authors in Major Cities

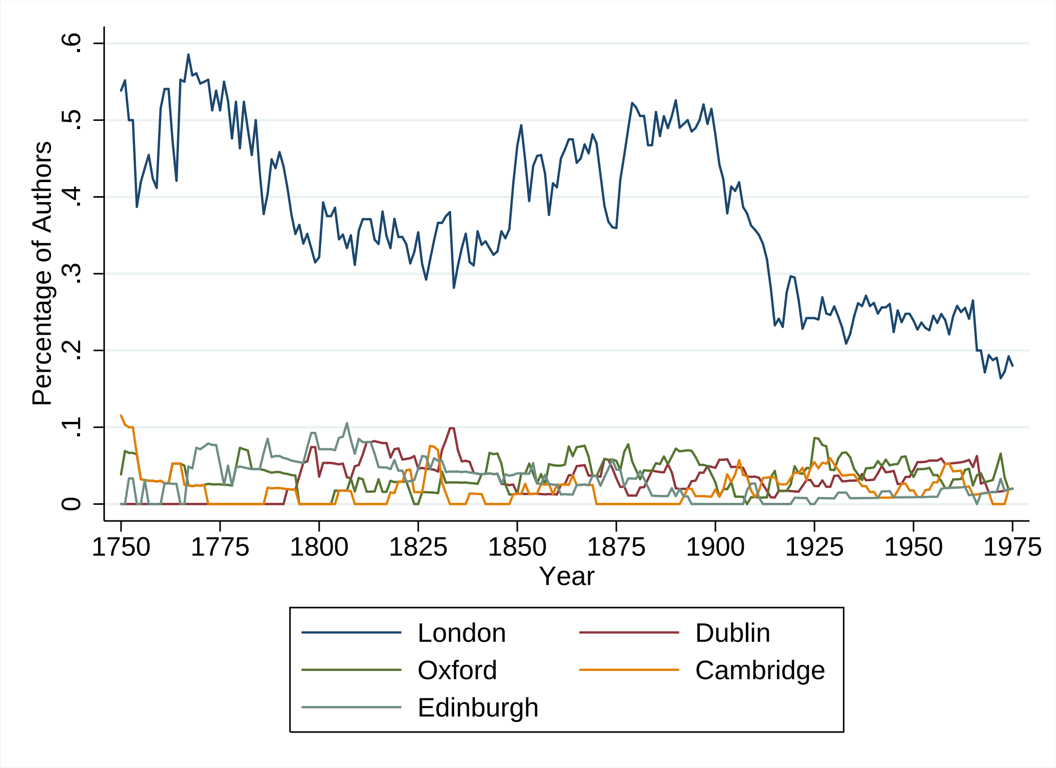

Figure 2: Total Authors and Output per Annum

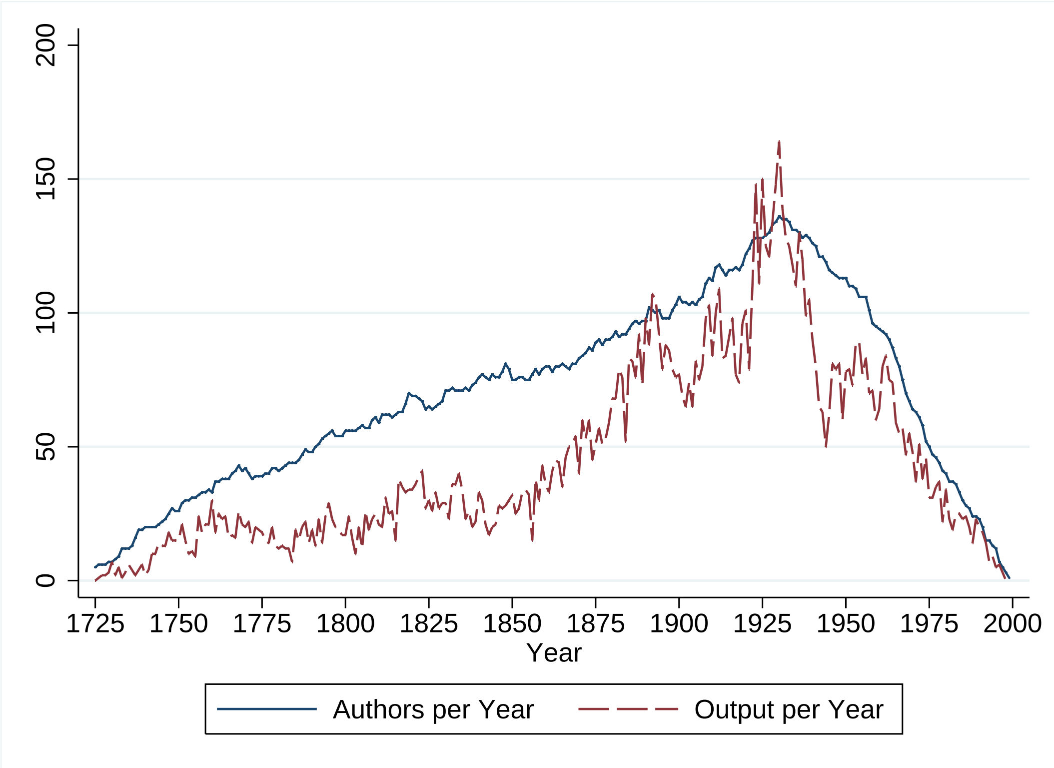

Figure 3: Total Output per Annum by Location

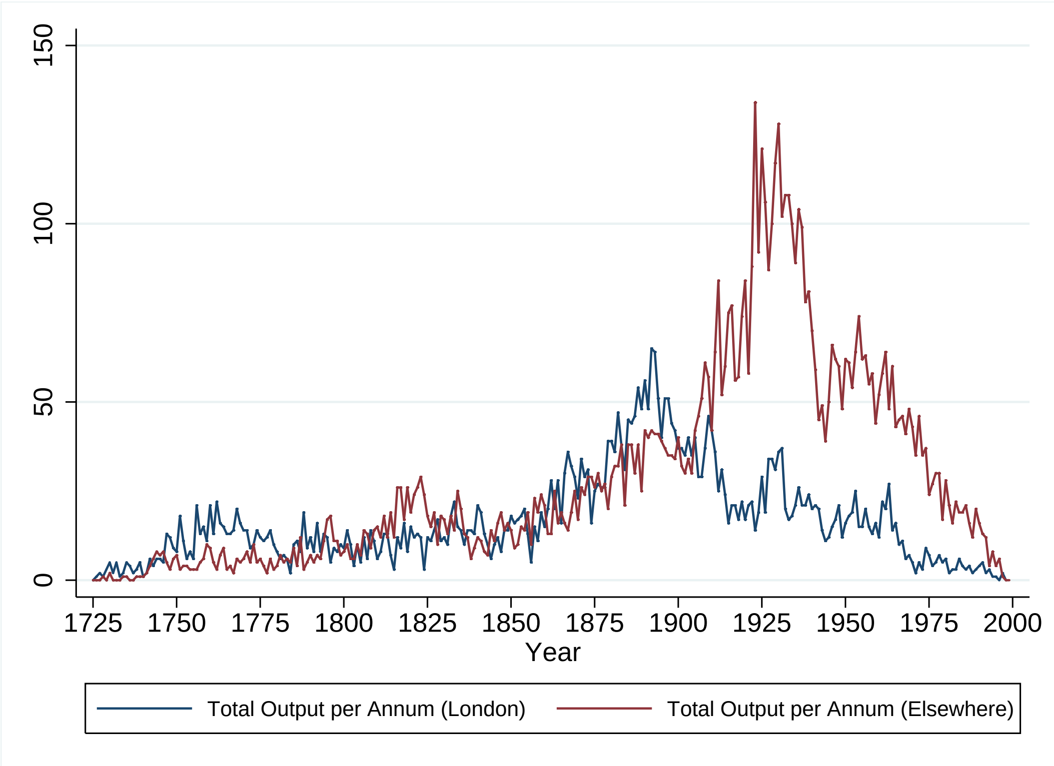

Figure 4: Frequency of Output Observations

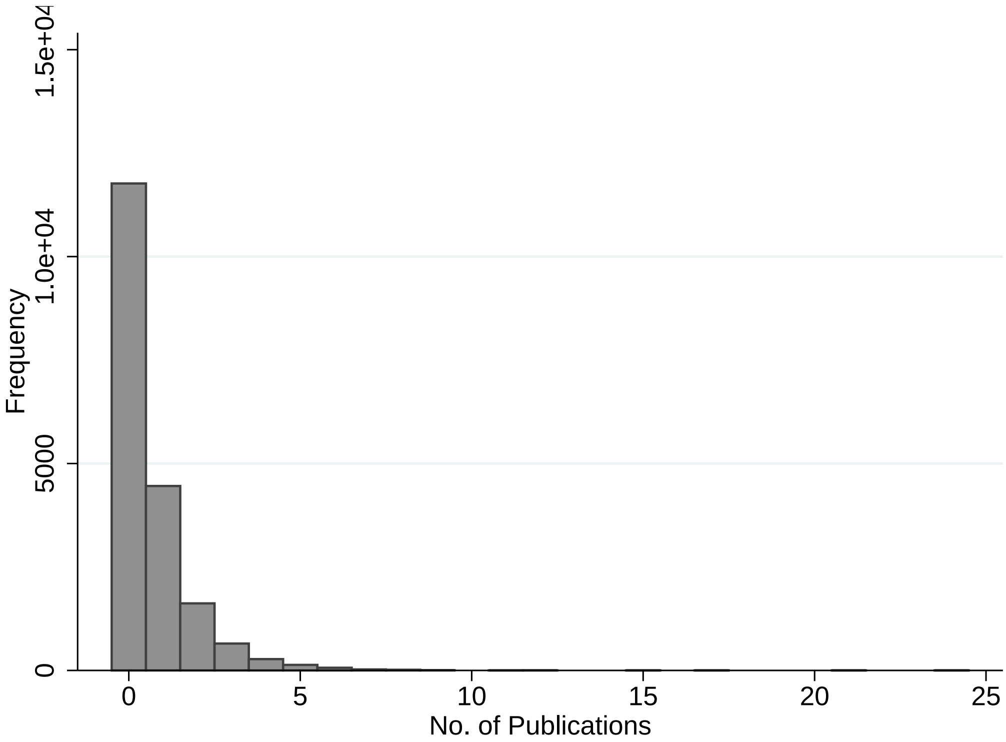

Figure 5: Impact of Static and Dynamic Effects on Age-Productivity Profiles

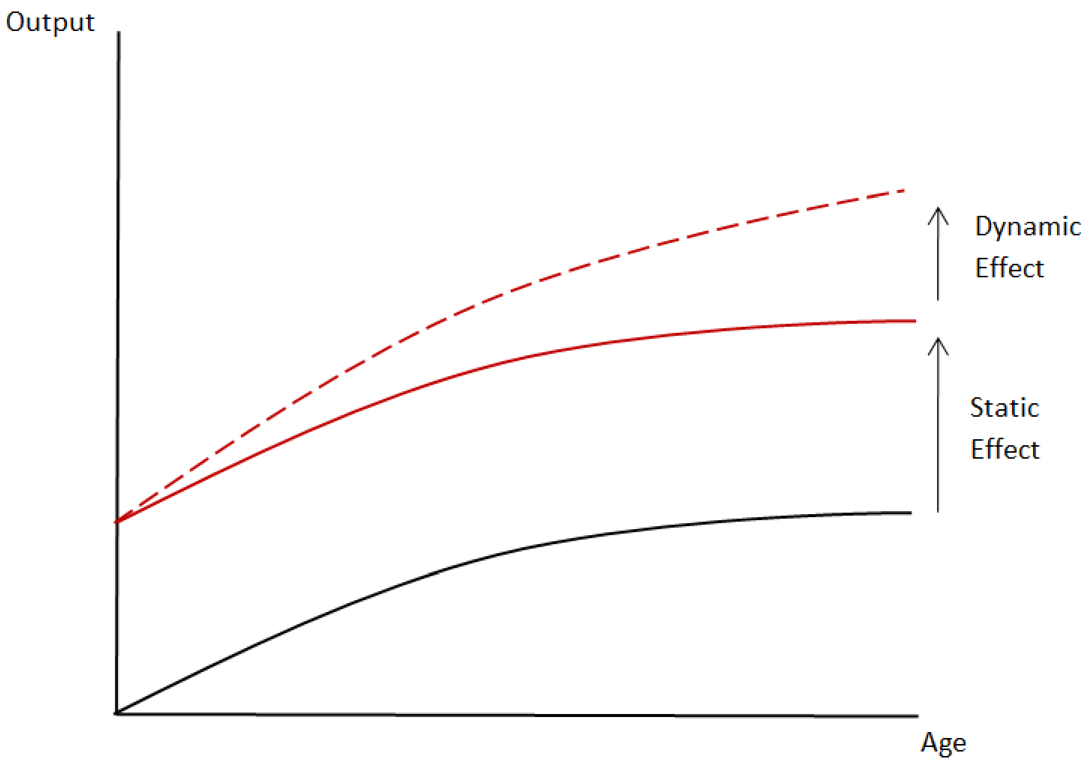

Figure 6: Estimated London Premium

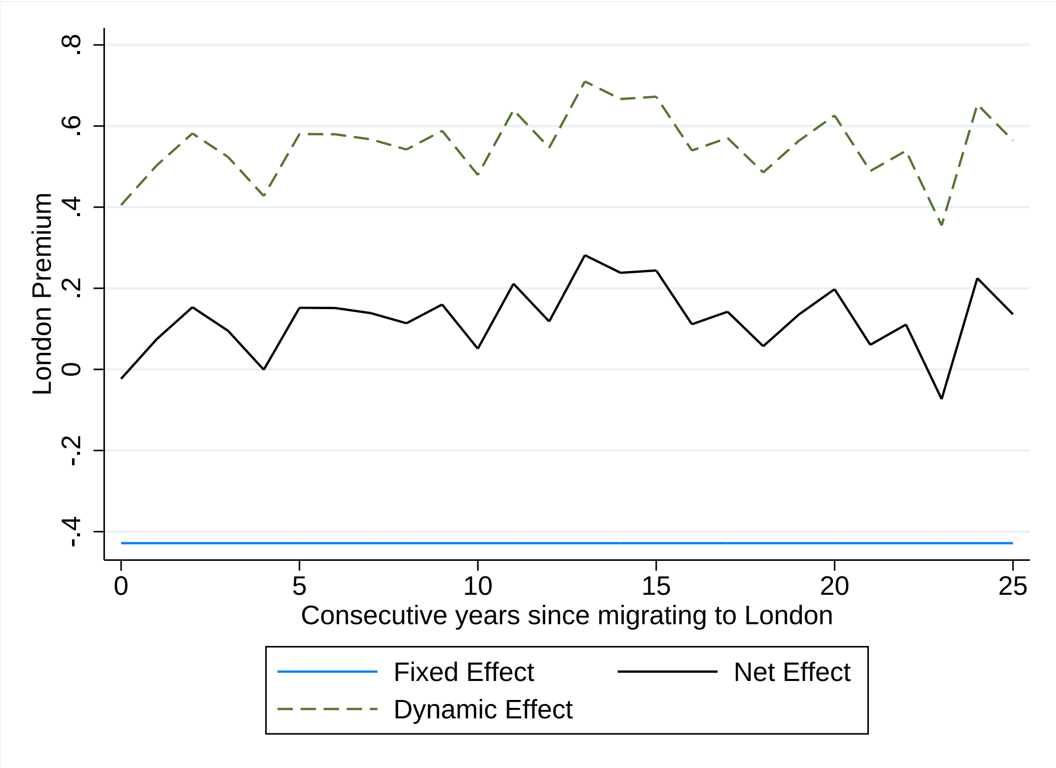

Figure 7: Estimated Pre- and Post- Migration Trend

B.2 Tables

Table 1: 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 Authors 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

Table 2: Summary Statistics by Gender

		Median	Mean	Std. Dev.
Male	Lifespan	69.00	65.96	16.43
N = 300	Age at First Publication	26.00	26.46	7.27
	Years in London	15.00	20.83	20.80
	Lifetime Works	23.00	34.70	32.44
	Career Length	41.00	39.50	17.23
	Works per Annum	0.67	0.92	0.79
Female	Lifespan	70.50	68.56	15.58
N = 70	Age at First Publication	27.50	28.33	8.06
	Years in London	21.50	24.29	20.51
	Lifetime Works	18.50	27.00	30.84
	Career Length	43.00	40.23	17.43
	Works per Annum	0.55	0.68	0.56
Total	Lifespan	69.00	66.45	16.28
N = 370	Age at First Publication	26.00	26.81	7.45
	Years in London	16.00	21.48	20.76
	Lifetime Works	21.00	33.24	32.24
	Career Length	42.00	39.64	17.24
	Works per Annum	0.64	0.88	0.76

Table 3: Top 30 Literary Artists by Ranking

Literary Artist	**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
Auden, W. H.	0.66	0.13	39.33
Orwell, George	0.64	0.13	38.26
Shaw, George Bernard	0.50	0.24	37.02
Eliot, George	0.42	0.31	36.28
Keats, John	0.49	0.23	35.89
Kipling, Rudyard	0.55	0.14	34.15
Shelley, Percy Bysshe	0.46	0.22	34.00
Moore, George	0.64	0.03	33.81
Browning, Robert	0.45	0.22	33.75
Carlyle, Thomas	0.55	0.11	33.05
Wilde, Oscar	0.43	0.22	32.50
Stevenson, Robert Louis	0.56	0.09	32.29
Betjeman, Sir John	0.63	0.00	31.55
Meredith, George	0.57	0.06	31.25
Alfred, Lord Tennyson	0.41	0.21	31.16

A note on interpreting these values: the highest-ranked author is James Joyce. 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. Dickens has a WCI value of 0.75 (rounded to two decimal places), indicating that Dicken's biographical entries contained 75% the number of words as those of Joyce. Dickens has a CI value of 0.76, indicating he received 76% the number of citations as Joyce. Thus, Dickens' Impact Index value is 75.47 – the average of his WCI and CI values normalised to 100.

Table 4: Main Results

	(1) OLS	(2) Fixed Effects	(3) Negative Binomial	(4) NB IRR
Age	0.0766***	0.0870***	0.149***	1.160***
	(0.00511)	(0.00549)	(0.00468)	(0.00542)
Age-Squared	-0.000728***	-0.000738***	-0.00157***	0.9984***
	(0.0000510)	(0.0000492)	(0.0000426)	(0.0000425)
London	0.0331	0.0944***	0.218***	1.243***
	(0.0528)	(0.0364)	(0.0322)	(0.0400)
Constant	-1.232***	-1.562***	-2.143***	0.117***
	(0.134)	(0.158)	(0.266)	(0.031)
Time Dummies	Yes	Yes	Yes	Yes
Author FE	No	Yes	Yes	Yes
R^2	0.103	0.120		
No. Authors	370	370	370	370
Observations	19022	19022	19022	19022

Robust standard errors are clustered on the author level and are reported in parentheses.

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

Table 5: Probability of Migrating to London

	(1) Output Last Year	(2) Total Output to Date
Age	0.968*	0.967*
	(0.0167)	(0.0166)
Age-Squared	1.000	1.000
	(0.000187)	(0.000183)
Output	0.993	0.997
	(0.0552)	(0.00603)
Time Dummies	Yes	Yes
Author FE	Yes	Yes
No. Authors	265	265
Observations	14040	13769

Logit odds ratios reported. Standard errors in parentheses.

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

Table 6: Probability of Migrating to London Based on Quality (High/Low)

	Low Quality		High Quality	
	(1)	(2)	(3)	(4)
	Output Last Year	Output To-Date	Output Last Year	Output To-Date
Age	0.948**	0.949*	0.982	0.982
	(0.0256)	(0.0255)	(0.0225)	(0.0226)
Age-Squared	1.000	1.000	1.000	1.000
	(0.000296)	(0.000293)	(0.000248)	(0.000241)
Output	1.066	1.002	0.944	0.992
	(0.0874)	(0.0102)	(0.0671)	(0.00809)
Time Dummies	Yes	Yes	Yes	Yes
Author FE	Yes	Yes	Yes	Yes
No. Authors	82	82	134	134
Observations	6986	6873	7054	6896

Logit odds ratios reported. Standard errors in parentheses.

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

Table 7: Probability of Migrating to London by Century of Birth

	18th Century		19th Century		20th Century	
	(1)	(2)	(3)	(4)	(5)	(6)
	Output Last Year	Output To-Date	Output Last Year	Output To-Date	Output Last Year	Output To-Date
Age	1.001	1.015	0.970	0.968	0.978	0.992
	(0.0371)	(0.0391)	(0.0238)	(0.0235)	(0.0416)	(0.0422)
Age-Squared	0.999	0.999	1.000	1.000	1.001	1.000
	(0.000426)	(0.000417)	(0.000275)	(0.000276)	(0.000455)	(0.000444)
Output	0.888	0.971*	1.013	1.005	1.144	0.992
	(0.105)	(0.0162)	(0.0704)	(0.00731)	(0.168)	(0.0162)
Time Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Author FE	Yes	Yes	Yes	Yes	Yes	Yes
No. Authors	75	75	150	149	82	82
Observations	3683	3683	8163	7892	2194	2194

Logit odds ratios reported. Standard errors in parentheses.

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

Table 8: Duration Effects

	(1)	(2)	(3)	(4)	(5)	(6)
	FE 1	FE 2	FE 3	NB 1	NB 2	NB 3
Experience	0.0276***	0.0265***	0.00639	0.0507***	0.0500***	0.00768*
	(0.00494)	(0.00568)	(0.00556)	(0.00339)	(0.00452)	(0.00464)
Experience-squared	-0.000683***	-0.000665***	-0.000185	-0.00117***	-0.00116***	-0.000234**
	(0.000115)	(0.000124)	(0.000117)	(0.0000820)	(0.0000967)	(0.0001000)
London		0.0159	0.0685*		0.0100	0.187***
		(0.0437)	(0.0405)		(0.0428)	(0.0434)
Age			0.0851***			0.147***
			(0.00546)			(0.00481)
Age-Squared			-0.000719***			-0.00154***
			(0.0000490)			(0.0000439)
Constant	-0.154	-0.159	-1.537***	0.551**	0.549**	-2.110***
	(0.120)	(0.121)	(0.157)	(0.220)	(0.220)	(0.268)
Time Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Author FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.074	0.074	0.121			
No. Authors	370	370	370	370	370	370
Observations	19024	19024	19022	19024	19024	19022

Robust standard errors are clustered on the author level and are reported in parentheses.

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