

Manhattan Transfer: Heterogeneous productivity effects of agglomeration in American authorship*

Lukas Kuld¹, Sara Mitchell², and Christiane Hellmanzik³

¹Department of Economics, Kemmy Business School, University of Limerick

²Department of Economics, University of Southern Denmark

³Department of Business and Economics, TU Dortmund

22 October 2024

We investigate quantity and quality effects of agglomeration in the careers of American authors. We combine novel yearly data on publications and work location of 471 eminent authors with US Census data to analyse industry concentration and agglomeration economies from 1850 to 2000. While finding a positive overall effect of living in New York City on the publication propensity of literary works, we focus on the heterogeneity of the effect along three axes: decade, age, and length of residency in NYC. First, the effect size correlates with industry concentration and maturity. Second, authors immediately increase publications after arriving in NYC, while the effect wanes after around 10-15 years. Third, the effect is strongest for younger authors in their 20s and 30s. In addition, works published while an author lives in New York City are more likely to achieve critical acclaim and to have lasting influence in terms of present-day popularity.

JEL-codes: N30 • N90 • O33 • R11 • Z11

Keywords: Agglomeration economies, urban history, geographic clustering, productivity, literature, creativity

*Corresponding author: Lukas Kuld, lukas.kuld@ul.ie. The authors thank Maria Fleming, Seán McGuane, Vivien Gaida, Leyla Ruzdija, and Jens Kirsten for their excellent research assistance. We thank John O'Hagan for his support and supervision of the first stage of the data collection process, and the Department of Economics at Trinity College Dublin for their financial support of the first stage of data collection. We also thank the participants of the North American Meeting of the Urban Economics Association 2021, the 2021/2022 FGCU, Portsmouth, Limerick, and Maynooth seminar series, and the 2019 European Workshop on Applied Cultural Economics for their comments on various versions of this paper. Finally, we thank Leah Brooks and two anonymous referees for the improvements made to the manuscript during the editing process at RSUE.

1 Introduction

In April 1900, Upton Sinclair fled the “ ‘madding crowd’s ignoble strife’ and deafening roar of the city’s commerce” [New York Times, 1901] to write *Springtime and Harvest* in a Quebec cabin. However, he returned to New York City after the summer to approach four different publishers before using the city’s printing and book binding infrastructure to self-publish the novel [New York Times, 1901]. This young author’s struggles and similar anecdotes hint at agglomeration effects in 1900 New York City when turning a manuscript into a publication. However, not only the American publishing industry and literary scene, but also communication and transport technologies have changed markedly over the last 150 years.

During the last decade of the 19th century, the American literary market entered a period of rapid transformation. Fast growing demand from an expanding base of educated readers, falling costs due to new printing technologies, and legal changes which eased the trade of literary texts induced the transformation of publishing into a modern industry during the first half of the 20th century. In this paper, we investigate the existence and magnitude of agglomeration economies in 150 years of American literary production observing both the rise and relative decline of New York as a literary cluster.¹ We show how New York City offered significant productivity benefits to authors residing there, most pronounced during the early stages of the industry’s professionalisation path. In turn, authors responded to these benefits by moving there.

To investigate industry concentration and agglomeration effects, we use data from historical US Censuses and a novel panel dataset which includes the complete careers of 473 prominent authors of American literature obtained from encyclopedias. We also combine several sources to measure the quantity and quality of publications produced by these authors. Due to our sample’s composition our estimates cannot be seen as the average treatment effect but the effect on a talented, relatively young subgroup of authors. However, these data allow us to draw a picture of the industrial development and evolution of agglomeration effects in American literary production from 1850 to 2000.

We estimate the difference of the publication propensity while being located in New York City and elsewhere for a working-age author to assess evidence for overall agglomeration economies,

¹In this paper, we use “New York” and “New York City” interchangeably to refer to the urban agglomeration around New York City. Based on the data, we define this agglomeration by a 30 km radius.

with a particular focus on understanding the heterogeneity in the effect by age, decade, and length of residency in New York City. Conditional on author, age, and time, we find that, on aggregate, an author has a five percentage point higher probability of publishing during the periods of her life when she lives in New York City than when she lives elsewhere. This finding, a 15 to 20 percent increase relative to publication averages, is in line with the literature. However, we find significant heterogeneity in this quantity effect.

First, inexperienced authors benefit more. The effect on publication propensity is strongest in the first 10-15 years after relocating to New York City and authors in their twenties and thirties benefit more than older authors. The immediate increase in publications after relocating to New York City suggests a significant role of factors lowering publication barriers such as matching with publishers or learning about the publication process.

Second, the publication differential between New York City and other locations rises with the establishment of a large commercial publishing market in the city around 1900. For long-term New Yorkers, however, the effect then declines with the city's industry share and the advanced professionalisation of the industry. Therefore, the importance of being located in New York City diverges between potentially more and less connected authors.

Third, works published while an author lived in New York City were more likely to achieve critical acclaim at time of publication and are more likely to have lasting influence in terms of present-day popularity. Localised benefits for contemporaneous commercial success, however, are less clear. Thus, co-location with other writers might be more important for success within literary circles and peers than with the wider public. Our identification largely exploits a staggered difference-in-difference setting using an author's move to New York City as an event study. Conditional on author, period, and age controls, we find no pre-trends in publications prior to an author's move to New York City in our sample. Moreover, our findings are robust to matching on gender, age and ethnicity.

A remaining concern is reverse causality through sorting into New York City of particularly high-profile or promising authors who would have also been successful in other locations. In our empirical set-up, such selection is partly addressed by a focus on publication patterns within an author's career through fixed effects which control for static differences such as an author's talent. However, we cannot rule out endogenous aspects of the timing of an author's move to

New York City. In fact, the jump in publication propensity directly after moving to New York City and anecdotal evidence, as described in the opening paragraph, suggest that an author with a promising manuscript might decide to move to New York City to advance their career. We cannot rule out that said manuscript would not have been published regardless of the author's location. This would bias our estimates, in particular, if a large share of authors misjudges the importance of location.

We base the causal interpretation of our estimates, that is the estimates reflect a true impact of location, on the following observations. The strength of agglomeration effects by decade corresponds widely to the effect size expected based on the state of the publishing environment in the United States and New York City during the time period. However, the age at which authors move to New York City remains stable over time, and we find no effect for moving to other large cities. Similar to a placebo test, we interpret age and large city moves as proxies for a writer's circumstances when moving to New York City. Therefore, the diverging estimates suggest that the relative publication effect following a move depends more on the state of the destination location and less on the writer's personal circumstances.

Moreover, the finding that established authors are the first to exploit opportunities in an emerging publishing market in New York City, but depend less on location during later decades, aligns with theoretical expectations as outlined in the next section. Finally, the findings on publication quality further support a causal interpretation of the observed publication differential as authors have less choice over the timing of their peak quality than their career more generally. Nonetheless, we acknowledge that there might be remaining issues - such as reverse causality - which we cannot control for with the data available to us. Importantly, while we dis-aggregate the effect by age, time period, and length of residency, we are not able to disentangle the mechanics of agglomeration effects such as matching and learning [Duranton and Puga, 2004].

The shown dynamics in the agglomeration of literary production and its heterogeneous effects on literary writers expand on studies that show (long-term) agglomeration benefits in the arts (e.g. Mitchell [2019]; Hellmanzik [2010]; Borowiecki [2013]; Tao et al. [2019] and Borowiecki and Dahl [2021]). On the one hand, we explicitly model and account for the heterogeneity in agglomeration benefits and provide both quantity and quality estimates. On the other hand, American census data enable us to use complete employment data for literature adjacent occu-

pations to draw a full picture of agglomeration dynamics, and without relying on the sample that is used to study agglomeration effects. *Dis*-aggregating the effect of living in a creative agglomeration over a long time period on individuals at different career stages and length of residency in the agglomeration (treatment length) alleviates two major concerns. First, the aggregation of heterogeneous effects with different cohorts and staggered entry can lead to a biased and even negative weighting of effects for sub-samples in a difference-in-differences setting. Second, the dynamics of the effect revealed along the three time dimensions hint at the mechanics of agglomeration benefits, while taking into account that the average New York City writer’s characteristics changed markedly over time. As described above, we find that the aggregate estimate is driven by subgroups of authors and, therefore, prone to misinterpretation without considering its heterogeneity.

Our findings further the understanding of long-term agglomeration effects and their dependence on market structure more generally. We thereby add to the rich economic literature on individual industries to understand agglomeration (e.g., Saxenian [1996], Arzaghi and Henderson [2008], Buenstorf and Klepper [2009], Cabral et al. [2018]) and, in particular, to the growing literature on the productivity effects of agglomeration in historical contexts (See Hanlon and Heblich [2021] for an overview). Time series data on firm activity are typically too short to assess agglomeration benefits in the medium to long-run and to observe the rise and fall of the industry. Our setting allows us to provide a dynamic assessment of agglomeration economies from a fledgling to a fully mature industry.

Moreover, our measure of productivity allows us to overcome a common limitation in the literature on long-run agglomeration economies (e.g., Hanlon and Miscio [2017]; Klein and Crafts [2020]). Direct measures of productivity are often not available in historical studies, so many studies rely on city size, employment levels, or patents as a measure of agglomeration forces [Hanlon and Heblich, 2021]. However, we are able to explore the impact of agglomeration on a direct measure of individual productivity (publications) rather than relying on a proxy measure.

We also extend the literature on innovations outside the patent system (e.g., Li et al. [2018], Giorcelli and Moser [2020], Moser [2012], Mitchell [2019], Hellmanzik [2010], Borowiecki [2013]), assuming that every book constitutes an innovation. Importantly, all publications in our data are commercialised and on the market, thus we only observe true innovative activity. In addition,

we provide several quality metrics spanning short- and long-run success, including measures of long-run critical acclaim (via inclusion in *Kindlers Literatur Lexikon*), contemporary commercial success (via inclusion in *Publishers Weekly* bestseller lists), and current day popularity (via the number of online ratings on *Goodreads.com*).²

Finally, we also contribute to a growing literature on the location of innovative workers and other significant figures across time and space, such as [Schich et al., 2014; Åke E. Andersson et al., 2014; Cabello and Rojas, 2016; Borowiecki and Dahl, 2021; Serafinelli and Tabellini, 2022], among others. However, many such studies only observe birth and death location, a common trade-off of using very large samples. Similarly, studies on long-run agglomeration effects and inventive activity tend to have relatively few observations per entity, for example, only identifying location at the time of patent filing or at the time of a census or survey or manufacturers. In contrast, our dataset includes author location on an annual basis across the entire lifecycle. We observe, for instance, that the common empirical restriction to only observe birth and death locations can hide substantial geographic concentration due to life-cycle migration. We can also directly compare our location patterns to the location of authors in the full count US census, which by definition does not document the childhood of future authors for instance. Importantly, the full migration history of authors allows the use of relocation to study productivity effects of agglomeration.

The rest of the paper is structured as follows. The next section provides an overview of the empirical setting and sets a conceptual framework for agglomeration economies in literary writing. Section 3 describes the datasets and the data collection methodology, and in Section 3.3 also provides an analysis of agglomeration patterns in American literature. In Section 4.1, we set up our strategy to identify individual agglomeration benefits. Results are discussed in Section 4.2. Finally, Section 5 concludes.

2 Empirical setting and conceptual framework

Our data include American authors either born between 1800 and 1949 and listed in the *Britannica* or registered in the US Census from 1850 forward and, therefore, allow us to capture

²For a descriptive overview of American authorship based on *Britannica* authors and authors in *Publishers Weekly* bestseller lists, see O’Hagan [2021] and Ceulemans et al. [2020], respectively.

a full historic picture of the industrial development of the publishing industry and the literary production associated with it. We will first provide the historic backdrop of the industry studied in this paper.

2.1 The development of the American book publishing industry

In the years after American independence, British literature remained central to American book culture, with most books in the US being imported from London, Dublin, and Edinburgh [Green, 2010] as reprints of massively popular and commercially proven British authors, such as Charles Dickens and Sir Walter Scott. In fact, publishing houses rushed to publish the newest novel as soon as copies arrived on ships from Britain.³ Due to its geographically convenient location, New York City became the leading port in terms of book imports and the centre of the publishing industry by the mid-1800s [Green, 2010; Glaeser, 2005].⁴ Around a third of the bookbinders in the United States were located in New York City in the mid-19th century, as seen in Figure 1. By 1860, New York City was the centre of the printing industry and had the largest concentration of writing-related occupations, such as editors and reporters.

The early development of the American book manufacturing and publishing industry was without much benefit to American literary authors [Green, 2010; West III, 1989] who “could be had cheap compared with paper, type, founders, and compositors” [Jaszi and Woodmansee, 2009, p. 91].⁵ ⁶ Therefore, authorship developed as an economically viable activity only after New York City had established itself as the centre of the printing and publishing industry. However,

³Until international copyright was recognised in 1891, publishing houses operated under an informally-enforced “courtesy of the trade” in which the first publishing house or printer to announce a foreign book, usually via newspaper, had the right to print it and no other printer would print the same book. For works by popular authors, publishing houses printed copies as quickly as possible without waiting to see who announced the book first [Green, 2010; West III, 1989].

⁴As Glaeser [2005] explains, New York City had a deep harbour that was free from sea ice and was relatively easy for ships to enter and leave, and New York City was close to Hudson River which served as an important distribution point inland. New York City was also a geographically convenient distribution point for the book market because it had relatively easy access to other sea ports in the United States. Later, the construction of the Erie Canal “not only opened an easy path to the upper part of the West from New York but helped make the land through which it passed one of the most active book markets in the nation” [Green, 2010, p. 119]. See Glaeser [2005] and Green [2010] more detailed discussions on the competitive advantages of New York City and their impact on both the city and the publishing industry.

⁵Green [2010] notes that only 20 novels of more than 100 pages written by American authors were published in the 1790s, and only 25 novels by American authors were published in the next decade. At that time, most of the novels read by Americans were imported volumes borrowed from circulating libraries.

⁶See Jaszi and Woodmansee [2009] and West III [1989] for a detailed discussion of the evolution of authorship in the United States during the 19th century and how it both impacted and was impacted by the development of copyright laws. Interestingly, American authors who were successful until the mid- to late-19th century often had individual or institutional patrons [Jaszi and Woodmansee, 2009; West III, 1989].

towards the end of the 19th century, the development of a highly concentrated and competitive printing and publishing industry facilitated the development of the magazine and advertising industries. These provided important income streams for authors [Williams, 2007; West III, 1989]. At the same time, there was a growing “cultural patriotism that encouraged Americans to become authors” [Williams, 2007, p. 116]. This shift occurred relatively quickly: the first list of bestselling novels released by Publishers Weekly in 1895 included only two American authors in the top 10 list; five years later, nine of the top ten novels were by American authors.

While New York remained the most important publishing location during the beginning of the 20th century, there have been several shocks which resulted in a decline in the concentration of the printing and publishing industries in New York City around the middle of the 20th century. Technological advancements in printing technology dramatically reduced the cost of printing, allowing for the expansion of smaller printing presses for books, magazine, newspapers, etc. across the United States [Winship, 2009].

The development of a national book trade system further facilitated a rapid expansion in the number of publications.⁷ Thus, the industry became bigger, more competitive and more dispersed across the United States and with that development it became easier to publish outside New York City.

A second large change for New York as the centre of publishing was the film industry beginning to emerge in Hollywood from the 1940s, expanding the number of professional opportunities for some writers outside New York City. Around the same time, there was an increased professionalisation of the publishing industry with agents becoming more important as brokers to publishing houses. These literary agents had become an established intermediary between authors and publishing houses due to the growth in subsidiary rights and more complex contracts [Kaestle and Radway, 2009; West III, 1989]. Thus, this period was characterised by a combination of factors that both decreased the advantages of being in New York City and decreased *disadvantages* of being located elsewhere.

Thirdly, World War II certainly changed the cultural landscapes of the late 30s and 40s in the US and particularly its larger cities. The productivity effects on the US of the European

⁷Winship [2009] notes that the US population more than doubled between 1880 and 1916 while the number of new book titles and editions increased sixfold. Luey [2009] observes that “title production increased doubly fivefold between the late 1940s and 1990” (p. 35).

exodus have been analysed for scientists (Waldinger [2012] and Moser et al. [2014]) and beyond the scope of World War II for innovation in the US due to migration by Diodato et al. [2021]. Particularly interesting for the purpose of our paper is Borowiecki and Dahl [2021]’s finding that there have been larger numbers of actors and actresses in New York City in particular.

And lastly and more generally, the rise of automobiles, road freight and telecommunication reduced the need for high-density cities. Indeed, all major cities experienced some degree of population decline during the period from the 1930s to the 1990s, although this decline was less dramatic in New York City compared to all other major cities except Los Angeles [Glaeser, 2005]. These technological shocks did not necessarily decrease the advantages of living in New York City so much as they decrease the *dis*advantages of being located elsewhere.

2.2 Agglomeration, industrial co-location, innovation, and technology in literary production in New York

A major reason for agglomeration is the physical proximity to other creatives in order to benefit from spillovers and to obtain access to the latest innovations. This was shown for the creative as well as scientific industries in [Borowiecki, 2013; Farrell, 2003; Waldinger, 2012; Moser et al., 2014] among others. Another obstacle to success for American writers was the difficulty of establishing meaningful social connections with publishers or others in critical gate-keeping positions, such as literary critics when working in the periphery. Moving to New York City would increase physical proximity to gatekeepers and therefore increase the probability of having the face-to-face interactions necessary for building trust, developing professional relationships, and ultimately getting published.

Before the rise of intermediaries, the author-publisher relationship was effectively a two-sided matching market⁸ and, in its taste aspect, was not too dissimilar to the Roth and Sotomayor [1992] marriage market model, with both also sides taking considerable financial risk. The pooling of authorship-related industries in New York, however, would have reduced search and opportunity costs [Helsley and Strange, 2002] in terms of writers’ ideas and resulted in higher

⁸West III [1989] describes the search process for a publisher: “What kind of publisher is best for a serious literary author? [...] Authors and publishers often come together by chance; sometimes an author is dissatisfied with a previous house, or the new house is rumoured to be the ‘right’ kind of firm. Most of the matches are imperfect, but occasionally an author does find an entirely compatible publisher.” (p. 25).

quality matches between authors and publishers [Helsley and Strange, 1990]. Thus, co-location to publishers in New York may have been particularly beneficial for early-stage authors with no track record. This is comparable to scientists without an established network of co-authors or early career workers benefiting from working at large multinational companies.

It is important to note that these type of agglomeration effects reflect the *relative* advantages associated with the concentration of economic activity in New York City, and therefore also the penalty associated with living elsewhere, for example, due to a lack of a local publisher. However, as the differences either decline or lose importance the return to authors to being in New York with its publishers will also wane over time. The reasons for this are threefold. Firstly and most generically, decreasing transactions costs due to better travel and telecommunication rendered the need for co-location less pressing in order to make it as a writer. Secondly, cheaper printing technology and the rise of literary agents as intermediaries decreased the penalty for authors located outside New York as publishing opportunities could be found elsewhere more easily. And thirdly, the competition from Los Angeles which offered many novel opportunities for writers, offered additional employment opportunities for authors [West III, 1989].

After all, these secondary employment options in the industry next to the creative activity of writing were also a driving force for the rise of New York. That is, the thick labour market would have reduced the risk of engaging in a writing career as authors could be employed in a variety of industries requiring writing skills. This would provide authors with more employment opportunities and also allow them to signal their quality to printing houses or gain popularity. Moreover, the favourable market in New York also could have facilitated better matches with employers who allowed authors to pursue multiple income streams or move from publishing in periodicals to negotiating royalty contracts [Williams, 2007; Haller and Heuermann, 2020].

In sum, we expect to see large growth in the concentration of authors in New York City around the turn of the 20th century, after the publishing, printing, and bookbinding industries in New York City were already established. We also expect to see agglomeration effects emerge from this time, while there was an almost complete lack of access to the market - and to other authors - for those located elsewhere. We expect to see a decay of these agglomeration effects from the 1930s when there is a notable decrease in the disadvantages of being located elsewhere, if agglomeration benefits are primarily driven by publication opportunities.

3 The data

This paper utilises several historic and contemporary data sources to identify trends in industry concentration and estimate agglomeration effects associated with New York City for US authors and the publishing industry. First, we use cross-sectional population data from historical US Censuses in order to identify trends in concentration of the book industry with a particular focus on New York. Second, we use a purpose-built panel dataset with lifetime data on biographies, places of residence and publication quantities for 473 prominent American authors. Moreover, we use three different measures of publication quality, namely current-day popularity via the number of ratings on *Goodreads.com*, critical acclaim via inclusion in the *Kindler Literatur Lexikon*, and commercial success via inclusion in *Publishers Weekly* top 10 bestseller lists. Thus, the unit of observation in our baseline estimations is the author whereas in the quality analysis it will be the publication.

3.1 Author and industry sample based on *US Census* data

We utilise *IPUMS USA full count Census* data from 1850-1940, 1% samples for 1950, two combined metro 1% samples for 1970, and 5% micro-samples for 1960 and 1980-2000.⁹ We identify a total of 47,619 responses for the 1950 basis occupation (OCC1950) category *006 Authors*.¹⁰ We examine high skilled occupations (professional and technical occupations as well as professors and instructors), which we define as the 1950 basis occupation category codes 000 to 099.

We also investigate several relevant industry occupations such as artists, editors and reporters, bookbinders, pressmen, printers, and apprentices in the printing trade.¹¹ For each group, we utilise data on total population, location, and demographic characteristics such as age, gender,

⁹More specifically, we utilise the Ruggles et al. [2020] IPUMS USA data. We use 1% microsamples for 1950 and 1970 because 5% samples are not available for these Census waves.

¹⁰Before 1960, the Censuses were conducted by US Marshalls or, from 1880, specially trained census-takers or enumerators via in-person interview [US Census Bureau, 2020]. The Census enumerators were instructed to ask for the interviewee to describe their occupation and ask follow-up questions for clarification as needed. The enumerator recorded the occupation as a string which was later classified into an occupational category. In the historic Censuses, the most frequent strings for *author* include: writer (543) author (496) authoress (173) waiter (29) poet (22) literature (21) literary (18) literary writer (13) auther (12) dramatist (11) autheress (9) playwright (9) story writer (9).

¹¹Artists are identified by the occupational code *004 Artists and art teachers*, editors and reporters are identified by *036 Editors and reporters*, and bookbinders are identified by *502 Bookbinders*. Occupations related to the publishing industry include *575 Pressmen and plate printers, printing* and *614 Apprentices, printing trades*.

ethnicity, location of birth (foreign-born or native-born), education level, and marital status. Importantly, Census data do not encompass information on literary production.

We use these historic US Census data to examine patterns of industry concentration and to substantiate New York’s status as a major literary cluster, comparing New York’s authors to those in other locations in terms of demographics. As such, the sample of Census authors serves as a baseline to determine whether prominent authors differ systematically from the average author population.

3.2 Prominent author sample, publication quantity and quality

The second set of data for this paper is a purpose-built panel encompassing eminent authors born in the US in the years between 1800 and 1949, their yearly literary output and three metrics to capture the quality of their yearly output.

Bibliometric sampling and productivity based on *Encyclopaedia Britannica*

We started by sampling all authors listed as American in the Encyclopaedia Britannica [2016] born between 1800 and 1949. Only those authors were considered who had at least one literary work, be it prose, poetry or a play. Thus, we exclude authors whose publications were limited to any other type of writing activity be it journalism, criticism, or song writing. For each author, we collected their year of birth, year of death, ethnicity, and gender, as well as annual data on the number of publications and residence from the authors’ biographies in Encyclopaedia Britannica [2016], The Literary Encyclopedia [2016], and Literature Online [2016]. The data on location were brought into a panel structure using the time of movement as the beginning and end point of any spell spent at a particular location.¹²

For these authors (henceforth referred to as the *Britannica* authors), we extracted information on the number of works published in each year and categorised them as novels, plays and poems. The encyclopedias often only included a list of “notable publications”, so the number of publications was collated from lists of publications across all three sources. To ensure completeness of our quantity measure, we added publications listed in *Goodreads*. While the names of

¹²Short-term travels were recorded; however, they are not considered in this study.

publications were not collected, the lists of publications were cross-referenced to ensure that publications were not double-counted.

This exercise totalled 4,258 publications over the years 1826 to 2013. Thus, the output measure indicates whether an author’s published work is listed in either the *Britannica* or *Goodreads* in any given year.¹³

Publication quality: *Goodreads.com*, *Publishers Weekly* and *Kindler* data

We analyse three different measures of a publication’s quality: present-day popularity, market success at the time of publication, and critical acclaim by contemporary experts.¹⁴

In order to obtain a measure of present-day market success, we retrieve ratings for the works of authors in our sample from Goodreads.com [2020]. Interestingly, ratings are only available for 352 of the 473 (or 74.4%) *Britannica* authors. For this sub-set, we retrieve data on 1,388 publications, including the first publication’s year and ratings on *Goodreads.com*. On *Goodreads.com*, users can list and discuss literature and rate each publication from one to five stars. We focus on how many ratings a publication receives as an indication of how widely known and read a publication currently is or, more precisely, was since the launch of *Goodreads.com* in 2006. As Bourrier and Thelwall [2020] note, the most popular books on *Goodreads.com* generally fall into two categories: recent ‘popular’ literature and older ‘classic’ (canonical) literature as, for instance, read in secondary schools. Therefore, present-day market success partly reflects critical acclaim and canonisation.

In order to assess a novel’s historical success, we collected the top ten bestselling novels in the United States in each given year from 1895 to 2011 as determined by *Publishers Weekly*. We scraped these *Publishers Weekly* lists of bestselling novels from Wikipedia, where they are organised by year [Wikipedia Contributors, 2019]. We then linked data on 164 bestselling novels to 79 authors in the *Britannica* dataset (equivalent to 16.7%). Thus, inclusion in the *Publishers*

¹³See O’Hagan [2021] for a descriptive analysis of the dataset, author migration patterns, and productivity over the lifecycle.

¹⁴It is important to note that we were not able to link publications based on title, as these were not collected from the *Britannica*. Therefore, when matching publications across datasets, we rely on the assumption that editorial teams of the respective encyclopedias and online databases maintain accurate records of the year of publication.

Weekly list reflects contemporaneous market success.¹⁵

Finally, we obtain data from *Kindlers Literatur Lexikon [2019]*, a comprehensive German encyclopedia on literature that aims to only cover the most important works in all languages in cultural history.¹⁶ As such, it is a valuable source of authors' best or canonical works as judged by the editorial team and therefore the measure represents experts' opinion.¹⁷ Ultimately, 226 publications included in the *Kindler encyclopedia* could be linked to 171 authors in the *Britannica* data set with around half of these authors only having one publication listed. As this variable is backward-looking in nature and not based on market success, we consider it a complementary measure of literary quality and reflective of a very long-run, ex-post quality assessment.

In total, the resulting panel data set includes 471 prominent US authors born between 1800 and 1949 and their 4,258 publications along with three different quality measures for sub-samples of their works. Summary statistics for the full dataset are provided in Table 1. This table includes the time period covered, total publications, range of publication years and the total number of identifiable and linkable authors included in each sub-sample.

3.3 Descriptive patterns of US historic literary production

Author location and the rise of New York as literary cluster

The results of mapping the places of residency of both the US Census as well as eminent authors (*Britannica* authors aged 18 to 65) across time are presented in Figure 2. Until 1990 New York is the single most important location for all authors with a peak in agglomeration of more than a third of authors in the first half of the 20th century. Consistent with the history and theory discussed in Section 2, most writing is scattered along the East coast with New England, Boston and Chicago in particular being important albeit much smaller centres. The West Coast, Los Angeles in particular, gains in importance as a place of residence for American writers from 1920 onward, while New York's importance wanes from 1950. Importantly, this decline in authors'

¹⁵For this particular empirical exercise the literary output considered are novels only.

¹⁶See Arnold, Heinz Ludwig [2009] for more information on the aims of the encyclopedia, the curatorial process, and the contributors.

¹⁷We received permission to collect data from the Kindlers Literatur Lexikon web portal through personal correspondence on 5 December 2017. The data collection process via webscraping was conducted in early 2018.

concentration in New York City is not due to a decline in authorship — the total number of authors in New York City continues to increase — but rather a marked increase in the number of authors located elsewhere.

Figure 2 shows a striking concentration of authors in New York City before World War II, especially when compared to the concentration of workers in other literature-related occupations such as bookbinders, publishers, and printers, editors and reporters, and other artists. As expected, the publishing and printing markets are established in New York City first, and then authors begin to concentrate there, in particular working-age *Britannica* authors.¹⁸

The concentration of *Britannica* authors across major literary clusters appears to be slower to respond to changes in the publishing industry than the overall author population. In 1850, Boston was the largest cluster of *Britannica* authors, even though the city had already waned in importance in the publishing industry and Census authors had largely relocated to New York City. For these prominent authors, New York City did not overtake Boston as the largest literary cluster until more than a decade later. By 1900, however, the concentration of *Britannica* authors in New York City surpassed that of Census authors. New York City's position as the largest cluster of prominent authors peaked around 1925 and remained steady until just before the 1950s, during which time 45-50% of all prominent authors lived in New York City. In contrast, the city's importance for the wider population of Census writers began to wane from 1930.

Geographical analysis of Census and *Britannica* authors

We compare the demographic characteristics of Census authors in New York City to authors living and working elsewhere in the US to determine whether there are systematic differences between these groups. The results of this exercise are presented in Figures 3 and 4.¹⁹ New York-based authors do not appear to be an exceptional group, in fact, for the majority of the period from 1850 to 2000, authors living in New York City are broadly similar to authors located

¹⁸Interestingly, the concentration of authors is most similar in its pattern to that of artists, although artists' concentration in New York City remained consistently lower than that of authors. The impact of the spatial constraints faced by artists in form of necessary access to physical capital (e.g., an art studio, canvasses, paintbrushes, paint, and other supplies) for spatial concentration, while not clear, could have suggested a higher concentration than authors.

¹⁹All summary statistics are calculated using the sample of working-aged authors (ages 18-65).

elsewhere with respect to mean age, the share of white authors, the share of female authors, the share of authors who never married (except after 1950), and the share of authors with at least some college education.

There are two notable exceptions. First, during the period of the highest concentration of authors in New York City (1900-1930), the average New York City author was around five years younger than the average author located elsewhere. Second, throughout the sample period, the average author in New York City is more likely to have been born abroad than authors located elsewhere, reflecting New York City’s history as a major hub of immigration.

Except for these two exceptions, the patterns are similar for the *Britannica* sample of eminent authors, as seen in Figure 4. For most of the time period, the *Britannica* authors residing in New York City and those living elsewhere are broadly similar in terms of age, ethnicity, gender, and having been born abroad.²⁰

Moreover, we explore the degree to which demographic and industry characteristics can explain author concentration using a prediction procedure. Specifically, we estimate the number of authors in a given county based on log industry employment in related industries, the log of total population, the respective shares of skilled workers, women, people born abroad and population that is white.²¹ We then use the parameters estimated to predict the expected number of authors in New York City based on the actual industry and demographic data of the seven New York agglomeration counties. Last, we divide the predicted values for New York City by the numbers predicted for all other counties to get the predicted New York share.²²

The results are illustrated in Figure 5. The actual share of authors located in New York City is higher than the predicted share but follows a similar trajectory.²³ Therefore, the declining author share after the 1920s mirrors the declining industry concentration and other socio-demographic factors. On the other hand, authors are even more concentrated in New York than these factors would predict at the very beginning of the twentieth century. Puga [2010]

²⁰Data on marital status and higher education are not available for the *Britannica* sample.

²¹We follow [Cameron and Trivedi, 2001] in using a negative binomial regression model as the dependent variable is overdispersed count data.

²²That is we first estimate $\log(\mathbb{E}(y_{i \notin NYC,t})) = x_{i \notin NYC,t}\beta$, with $x_{i \notin NYC,t}$ the vector of covariates for county i (outside NYC) in year t , and $y_{i,t}$ its number of authors. We then use the estimated $\hat{\beta}$ vector to predict the share of authors in the New York City agglomeration: $\sum_{i \in NYC} \hat{y}_{i,t} = \sum_{i \in NYC} \exp(X_{it}\hat{\beta})$. Last, we calculate the predicted share as $\sum_{i \in NYC} \hat{y}_{i,t} / \sum_{i \notin NYC} \hat{y}_{i,t}$.

²³The level of industry concentration is somewhat sensitive to the prediction method.

notes that “substantial localisation or spatial concentration of economic activity may be seen as a sign of agglomeration economies” (p. 204) as per the Starrett [1978] spatial impossibility theory.

Lastly, we analyse mobility patterns. Authors who move to New York City tend to do so before age 35, as illustrated in Figure 6 (histogram on right y-axis). This is the typical age range for labour migration (see e.g. Stark and Bloom [1985]), suggesting that authors move to New York City for professional reasons and market opportunities as consistent with human capital theory’s implication of higher pay-offs and lower relocation costs for young authors [Becker, 1964]. This is underpinned by the fact that the average age at which *Britannica* authors move to New York City remains relatively constant for most of the sample period (local average line in Figure 6).

The frequency of moves shows that a large part of the concentration of authors in New York City is due to in-migration and not due to a large share of authors being born in the city. The number of author relocations to New York over time closely aligns with the industry’s trend in concentration, that is, the number of moves increases until the 1930s and decreases thereafter (histogram on top x-axis of Figure 6). At no point in time would New York City have been the observed literary cluster purely with native New Yorkers.

4 Agglomeration benefits by decade, age, and length of residency

4.1 Empirical Strategy

As our baseline, we determine the existence and magnitude of agglomeration effects by estimating the difference in publication propensity while being located in New York City and elsewhere for our sample of working-age authors. Specifically, we estimate aggregates of β_{it} in the following equation:

$$\mathbb{E}(y_{it} | \cdot) = \beta_{it}NYC_{it} + \gamma_i + \tau_t + \alpha_{it} \quad (1)$$

where y_{it} is a binary indicator for a publication by author i in year t , NYC_{it} is a dummy for whether author i lives in New York City in year t and β_{it} the associated NYC effect for this writer and year. γ_i and τ_t are fixed effects for author and time period, and α_{it} is a fixed effect for the age of author i in year t .

To allow for heterogeneous effects, we follow Gardner [2022] in our main specification using Butts [2021]. This is akin to a two-stage process where we first use all author-year observations outside New York City to estimate the publication propensity based on author, age, and period fixed effects. That is, we first estimate $\hat{y}_{it} = \gamma_i + \tau_t + \alpha_{it}$ excluding observations of author i in year t if NYC_{it} equals one. In the second step, we estimate the residuals - or adjusted outcomes - $y_{it} - \hat{y}_{it}$ as a function of different NYC aggregates. For instance,

$$y_{it} - \hat{y}_{it} = \beta NYC_{it} + u_{it} \quad (2)$$

gives us the overall differential in publication propensity β , or, using dummies for being in New York City in a specific decade, we can estimate the publication differential specific to that decade k :

$$y_{it} - \hat{y}_{it} = \beta^k NYC_{it}^k + u_{it}. \quad (3)$$

Equation 1 is a generalisation of the specifications in the related literature that would typically use a more restrictive two-way fixed effects model or polynomial specifications to condition the estimates on author, age, and time (e.g. Mitchell [2019] and Borowiecki and Dahl [2021]). The approach also links with other studies on agglomeration economies with the difference that local wages are often used as a proxy for productivity under the assumption that workers are paid the value of their marginal product in competitive labour markets (See Combes and Gobillon [2015] and Ciccone and Hall [1996], among others). However, this empirical setting provides the unique opportunity to utilise a direct measure of individual productivity rather than relying on wage proxies.²⁴ These aggregated results are also directly comparable with the estimates in

²⁴This approach also relates to the literature on agglomeration in innovation, which uses patents as a direct measure of individual productivity. (See Carlino and Kerr [2015] for an overview.) The literary market is characterised by highly differentiated products, each of which can be considered an incremental innovation. The use of literary publications as a measure of innovation has a unique advantage over traditional measures of innovation such as patents: many inventions are patented but relatively few are commercialised and introduced into the market; whereas, we only observe publications that are commercialised and introduced into the market.

Hellmanzik [2010]; Borowiecki [2015]; Mitchell [2019].

Identification and heterogeneity of effect

In our empirical framework, identification of the New York City coefficient (an aggregate of β_{it}) in Equation 1 comes from individual authors' moves to New York. These moves are unlikely to be exogenous to an author's time-invariant characteristics (e.g., natural ability, socio-economic background) or time-variant characteristics (e.g, career stage, marital status).²⁵ Therefore, we explore whether the estimated effect reflects a causal relationship from being in New York to increased productivity as measured by the probability of publishing or whether it is in fact the result of confounding factors, such as social reasons which have successful authors move to New York.

In Equation 1, we condition our estimate on the individual author, time, and age. Therefore, we address constant productivity differences, be it between authors, between different ages or a general publication slump, for instance, during the Great Depression. Thus, identification is not threatened by the possibility that New York City-based authors are constantly more (or less) productive than authors based elsewhere.

Instead, our main concern is that the decision to move to New York is correlated with career stage in a way that is not captured by age fixed effects (e.g., if authors begin their career at different ages). We provide counterfactual estimates to investigate the causality of productivity changes after an author moves to New York City based on three 'non-treatment' scenarios: the authors' productivity before relocation, their peers' productivity elsewhere, and New York moves at times when industry concentration and maturity predict lower agglomeration economies.

First, we set up the move of working-age authors to New York as an event study. This allows us to explore their productivity trajectory before and after relocation. In addition, we can compare the productivity pattern to the productivity of their peers who never move to New York as a comparison group via a matching procedure. If the conditional (age, author, time) productivity levels after moving to New York City are an expression of the author's career that would have played out in a similar way without any New York agglomeration economies, we expect to

²⁵See for instance Baum-Snow and Ferreira [2015] and Combes and Gobillon [2015].

observe an acceleration in publications before moving, especially in comparison to peers located elsewhere. Otherwise, the move can still reflect a career step but the agglomeration economies in New York might have been a necessary condition for the career's success.

Second, we utilise our long time horizon and observed shift in New York City's dominance. If authors move to New York at a similar career stage - and possibly with similar motivations - throughout our sample period but their observed gains change in a way consistent with the expected agglomeration economies associated with industry concentration and maturity, then we can more confidently attribute the observed productivity gains to being in New York instead of an expression of factors correlated with the decision to move.

In the last part of our empirical analysis, we provide evidence on the effect on quality. This approach addresses within-author endogeneity, as authors arguably have less choice over the timing of their peak quality than when to start their career. If the observed productivity gains were driven by the fact that moving to New York purely coincides with the decision to start a career, it is unlikely that we would also observe these patterns of quality gains. Therefore, we can more confidently attribute these productivity and quality gains to agglomeration economies in New York City.

4.2 Findings

4.3 New York City agglomeration effect

We first show estimates for the overall differential in the propensity to publish a literary work for authors in and outside New York City conditional on time, author, and age fixed effects. Table 2 shows separate estimates for the publication differential overall, for the first ten years in the city, and after more than twenty years. In each scenario, the counterfactual publication propensity is constructed using authors who never lived in New York and the movers' years before relocating to the city.²⁶ The estimation is then repeated using different estimation approaches.

Column (1) in Table 2 displays our preferred specification, the publication propensity conditional

²⁶The estimations in Table 2 are based on around 11,400 yearly observations from 301 working age authors. We exclude authors who live in New York before 23 or move there after 61, and all observations after an author moves away from New York. This gives us 2269 observations in New York overall, 938 in the first ten years in the city, and 710 after twenty years. However, the last number is based on only 57 authors, the sub-group of authors who moved to New York as young adults and stayed long-term.

on author, decade, and age fixed effects using Gardner [2022] (See Equation 2 in the last section). In this setting, living in New York City is associated with a five percentage point increase in the publication propensity, that is, an increase of around one quarter (0.22) of the average publication propensity outside New York City .

When estimating the effect separately for the first ten years after moving, we see the effect is stronger for authors who just moved to the city. Conversely, we do not find an effect after twenty years of living in the city when omitting the first twenty years in the estimation. We will investigate this pattern in more detail below. However, the benefits of living in New York City are highly heterogeneous across years located in the city and mostly accrue in the first years after moving there.

The point estimates obtained are similar for alternative estimation approaches including Callaway and Sant'Anna [2021] without time fixed effects or using Gardner [2022] with year instead of decade fixed effects. A traditional fixed effect approach using OLS or logit gives conceptually equivalent estimates, though higher overall. As expected, the linear OLS estimates are very similar to the Gardner [2022] estimates for more homogeneous subgroups, that is, the first ten and over twenty years in New York City.

As a first robustness check of the overall New York City estimate, we investigate whether this differential purely reflects correlated effects associated with recent moving activity or living in a major city. If authors maximise their publication output and if, in addition, any move between locations inflicts further costs, such as social adaption, insecurity, and direct financial costs, it is reasonable to expect a positive benefit of any move on productivity. In Column (4), we repeat the estimates with controls, including an indicator for whether the author has moved to a new place in the last five years, the cumulative number of moves up to the current year, and whether the author has at least one previous publication. We find that the inclusion of these additional mobility-specific variables does not change our findings significantly, though we do find that authors with a high number of moves given their age also have a higher probability of publishing.

These findings are consistent with the findings of other studies on agglomeration economies in cultural production. Hellmanzik [2010] finds evidence of substantial premiums for paintings produced in the artistic clusters of Paris and New York. Borowiecki [2013] and Borowiecki [2015]

observe agglomeration economies in classical music production. More closely related, Mitchell [2019] finds that there is a significant productivity gain for London-based writers. Consistent with findings here, Mitchell [2019] does not observe evidence of agglomeration effects associated with living in minor literary clusters or cities.

The equivalent estimates for cities other than New York City are shown in Table 3 (Los Angeles, Chicago, San Francisco, Washington DC, Boston, and Philadelphia). The estimated publication premium in these cities is much lower and insignificant over all years (Column (1)) and also in the first ten years after the relocation (Column(2)).

So far, we have excluded observations from authors who moved away from New York City as a comparison group. In Column (3) of Table 3, we see that the publication bonus is similar in size while living in New York City and after leaving the city. Therefore, we do not see a large moving-away penalty for authors. We interpret this as an indication that agglomeration benefits are partly transferable. However, since these authors have made the decision to move, they might be less affected than the overall NYC author population. Therefore, similar to the effect of moving into the city, we interpret these estimates as being specific to this sub-group and not as a general treatment effect of moving away. On the other hand, given the costs of moving noted before, even a low penalty of moving away, might, therefore indicate more significant location effects in New York. This would be consistent with de la Roca and Puga [2017] who find persistently higher earnings for Spanish workers who had once been located in a big city.

Last, the overall effect is higher for authors in their twenties and thirties as shown in Figure 7. For comparability, the observations are restricted to the first ten years in New York City. As few authors move to New York in their fifties, the precision of the estimated effect is low. However, we find a similarly low and insignificant point estimate for authors in their forties. Since the age of movers is stable over time (Figure 6), these findings are further evidence that more experienced authors benefit less from living in New York City.

4.4 Moving to New York City as an event study

Since identification comes from moves to New York City, we zoom in on the yearly productivity of authors who move to New York in order to assess the trend in productivity around the time

of their move to New York. This helps our understanding of the agglomeration mechanisms at work as well as possible selection mechanisms. We consider all moves of working age authors to New York as an event study; that is, we analyse the productivity of these authors around their relocation to New York in two frameworks: a panel regression framework with yearly dummies relative to the year of relocation, and a matching framework for a comparison between New York-headed authors with similar authors who never move to New York.

4.4.1 Publication differential pre- and post-move to New York City

In Figure 8, we show the estimates of relocating to New York City extending the regression framework as described in Section 4.1 using dummies indicating the relative time to relocation to New York. On the left, we show the yearly difference in publication propensity, taking the authors' last year before the move as the reference period.

We see that in almost all years after the move to New York authors are more likely to publish conditional on author, time, and decade effects. Crucially, we do not observe an upward trend in the publication odds over the last five years before the move, but a clear jump in productivity immediately after the move to New York which is sustained through several years after the initial relocation with two major caveats. First, the productivity bonus declines after around ten years to around zero after twenty years - perhaps after the initial benefit has been fully capitalised on. Second, the yearly comparison gives us point estimates with considerable variance.

We, therefore, focus the interpretation on the aggregate development of the estimated effect over several years, and also explicitly model this using five-year periods in the right plot in Figure 8. We observe that the publication propensity is significantly higher for the first 15 years after relocation with an increased publication propensity of up to 10 percentage points (40-50% of the average outside NYC). On the other hand, authors are similarly productive during the five years before relocation compared to years before, and also after twenty years in New York City.

As described, the primary aim of this exercise is to determine whether there is a distinct jump or kink in the productivity trajectory after the move (and thus the gains can be more confidently attributed to residence in New York) or whether authors accelerated their productivity in the years immediately before the move to New York (suggesting that the estimated gains may

reflect correlated effects). However, this exercise also allows us to analyse the dynamics of the agglomeration effect. Since, we find that the gains remain relatively constant for the first several years after relocation and then begin to decay ten or more years after moving to New York City, the agglomeration effect on quantity seems primarily driven by static agglomeration economies. That is, we observe evidence that productivity gains associated with exposure to agglomeration forces can be accessed and exploited immediately. This finding is comparable to de la Roca and Puga [2017] results for Spanish workers' earnings in terms of the immediate static effect after relocation as well as a learning effect accruing over time. The fact, that dynamic effect for our sample of American star authors is different from the dynamic effects for Spanish workers might be due to the difference in sample or indeed the country or time period analysed. Lastly, the positive productivity effect immediately after moving to a cluster is similar to that found in Mitchell [2019] for London-based authors.

The channels through which the productivity gain materialises are manifold. Authors could learn in ways which impact the type of work they produce such as transitioning from short stories to novels or plays; the quality of their work could be increasing as indicated by writing a bestseller or critically acclaimed work. It is also possible that moving to New York City helped authors match with publishers who understand and respect the authors' artistic perspectives or offer a multi-book contract. It was not necessary to live in New York City to have access to the literary market or to find a publisher, but living in New York City and having access to a dense network of publishers likely made it easier. Once a good match has been made, an author's productivity might increase as the author no longer has to search for a new publisher for each respective publication therefore reducing search costs and increasing time available for writing. The rate of production will increase after this match has been made, but the rate of production would not necessarily continue to increase over time, especially if there are additional limits to publication quantity, such as author-side limitations in the writing process. Moreover, for the case of playwriting one can easily imagine that adaption opportunities were higher in New York City.

4.4.2 Matching of movers and non-movers

We repeat the analysis of the relative impact of relocation using a direct comparison of publication means. We compare the mean productivity of authors who move to New York City with authors who never live in New York. To this end, we match authors who move to New York between 20 and 60 with authors who never live in New York by exact age, year (± 1), ethnicity, and gender. Figure 9 illustrates the results.

We see that authors who move to New York do so on average at an age corresponding to their, and their comparison group's, career start. Both groups have a clear upwards trajectory in their yearly publication propensity. However, the New York *movers* group shows significantly higher yearly publication means than their comparison group during most of the early years in New York but not before their move there. Before, and long after the move, the difference between both groups is insignificant or, in fact, the comparison group's average is higher.

These findings provide further evidence that the observed New York City effect presented above is not driven by model choice. In addition, because we match authors by age and time, the effect is also not driven by some type of sample imbalance. For example, there are relatively few authors in New York City after World War II, but they are more productive than earlier cohorts even at later stages of their careers. Last, we do not observe evidence that highly prolific authors systematically self-select into New York City.

4.5 Agglomeration effects over time and by industry concentration

New York City's industrial dominance changed dramatically over the sample period. We link these dynamics in this section to agglomeration benefits directly and via decade-level estimates for the period 1880-1980 in Figure 10. This figure illustrates the estimated decade-specific NYC effect relative to the overall publication propensity. We obtain the estimates by dividing the publication differences obtained from Equation 1 in Section 4.1 for each decade (β^k) by the average publication propensity of authors during the respective decade (pr^k). We use β^k/pr^k to account for the rise in publication propensity over time as plotted in Figure 11. The 95% confidence intervals are obtained by repeating the estimation 500 times based on re-sampling the authors with replacement.

As in the main estimates above, we differentiate the NYC effect by the length of time a writer lives in New York to account for the heterogeneity in the sample composition and effect over length of time in New York. Figure 6 and 12 show the age at relocation to New York City and the distribution of years spent there over time. While the average age when moving to New York appears stable over time (Figure 6), the increased stock of authors who stay in the city pushes the average time of having lived there upwards (Figure 12). For instance, the median author in New York City has been there for 15 years in the 1930s, while the median 1920s author had lived there for around 10 years only. Therefore, we restrict the estimation to either (a) the first ten years or (b) authors who have been in the city for over twenty years to distinguish decade effects from length of treatment (time in NYC) effects.

We see three major patterns in the effect by decade as shown in Figure 10. First, we see no clear effect for writers before the 1890s. Second, the new-in-New York effect seems to fluctuate at around 30 percent of the average publication propensity from 1900 onward. Third, long-time New Yorkers benefit from their location only around 1900.

The first three decades of the twentieth century were the climax of industry transformation and author concentration in New York City (see Figure 2). We use these years as a reference point to understand the relative effect size during other decades in Figure 13. Based on 95% confidence intervals, we find significantly lower benefits for long-time New Yorkers (over twenty years) during most later decades. Some of these long-time New Yorkers might have benefited earlier, during the 1890s. The effect for writers during the first ten years in New York City is significantly lower before 1900, but not significantly different to this reference period during later decades.

We conclude that the rise in the New York effect coincides with the establishment of a professional publication industry in the city. Young authors who move to the city, seem to be able to increase their output after relocating once the publishing industry has been established there. Most interestingly though, authors who live in New York for over twenty years, that is more established writers, show an increased publication propensity (compared to their contemporaries elsewhere and their own productivity before moving) only during the early years of a dominating New York publishing industry. These years would have brought new publication opportunities to writers already living there and thereby resemble an agglomeration move, even

though the writers did not relocate. On the other hand, once a fully professional publication industry had been established in the 1920s, experienced writers could avail of services such as literary agents or use established links with publishing houses to publish their work.

With important caveats on sample imbalance and estimation issues discussed above, we also directly estimate the occurrence of individual productivity gains and industry concentration in New York City. We interact the New York City effect with either the share of authors in New York in a given year (Figure 14a) or the difference of the share over the last five years (Figure 14b) in a fixed effect logit regression. We observe that the agglomeration effect on productivity is higher at times of high industry concentration, when many authors live in New York City compared to the rest of the US. The second specification shows that the effect is higher when industry concentration increases, that is, when many authors recently moved to New York City. We compute industry concentration using US census data, so this statistic is computationally independent of the behaviour of the authors in the *Britannica* sample.

In sum, Figure 14 suggests that productivity gains associated with residing in New York City rise and fall in tandem with shifts in New York City's industry concentration over time. Given that the average age of moves to and from New York remains constant over time (see Figure 6), we can more confidently attribute the estimated productivity gains to changing agglomeration patterns in New York City as opposed to confounding effects due to the decision to start a career.

4.6 Evidence on commercial and critical success

It may have been more important for authors pursuing critical acclaim to be located in a literary cluster than for those pursuing commercial success. Literary historians often make the distinction between the anti-popular and original *romantic genius* who is in conflict with the market and the *popular writer* who caters to the commercial market [Williams, 2007]. Bourdieu [1985] describes these as the *field of large-scale cultural production* (the creation of cultural goods for non-producers of cultural goods or 'the public at large') and the *field of restricted production* (cultural goods created for an audience of other producers of cultural goods). He notes that the *field of restricted production* develops its own evaluation criteria that breaks away from that of

the general public.²⁷ Therefore, authors seeking critical acclaim may have benefited from the geographic clustering of literary and artistic elites by learning about up and coming literary styles through a network of peer writers. This environment may not have been as beneficial for authors pursuing commercial success because, by design, the tastes and preferences of a highly localised literary elite may not have been appealing to the masses.

Thus, we examine the effect of moving to New York City on success of published works with readership and literary critics. This is interesting to better understand the mechanisms involved but also as a check on our quantity estimates presented earlier. Authors can determine when to start their career but have little or no control over the timing of their peak quality. Table 4 shows estimates for the probability of publishing critically acclaimed works, included in the *Kindler* encyclopedia, and commercially successful publications, included in the *Publishers Weekly* bestselling lists. Column (1) estimates the probability of publishing a work which entered the *Kindler* encyclopedia. Column (2) shows estimates for having a book in the *Publishers Weekly* top ten in year t .

Using the same estimation process as before (based on Gardner [2022]), the results show that moving to New York City has a positive and statistically significant association with the probability of publishing critically acclaimed works. At around 100 percent, the gains for critically acclaimed works (3.45 percentage points) are greater than for publications overall when compared to the publication propensity for such works outside of New York City (3.7 percentage points). In contrast, authors are not significantly more likely to experience commercial success when living in New York City compared to living elsewhere (Column (2)).

These findings with respect to quality support theoretical arguments for cluster benefits: authors pursuing critical acclaim will receive greater gains from agglomeration than those pursuing mass market success. These findings could indicate that living in New York facilitates better matches between authors and relevant industry gatekeepers. They could also indicate that the transmission of tacit knowledge of literary styles, norms, and evaluation criteria was necessary to achieve critical acclaim. However, the quality effects in Columns (1) and (2) could partly be driven by quantity effects as authors with more publications would also be more likely to have more entries in best-of lists. Therefore, we next measure differences in the reception between

²⁷See also Bourdieu [1969], Becker [1974], among others.

publications instead of the yearly output.

Specifically, we explore the longevity or lasting impact of works that were published after an author moved to New York as measured by contemporary popularity via the number of ratings on *Goodreads*. Here, we use the term “contemporary” to mean present-day popularity, not contemporaneous (at time of publication) popularity. In Columns (3) and (4), we use all books listed on Goodreads to estimate the probability of receiving more than the median number of reviews (Column (3)) or the total number of reviews in logs (in Column (4)). We find that authors who publish while living in New York are more likely to publish a book with a high degree of contemporary popularity (i.e. more than the median number of ratings on *Goodreads*). In addition, the number of total ratings per book is also significantly higher after a move to New York City.

Therefore, these findings show the importance of having lived or living in New York City on contemporary popularity, that is, agglomeration effects which materialise over the very long run. As Bourrier and Thelwall [2020] show, ratings on Goodreads for older publications mirror literature read in secondary and tertiary education. The impact measured here is, therefore, probably at least partly transmitted by critical acclaim. As such, these findings complement our previous findings on critical acclaim using the *Kindler* encyclopedia, albeit with a more indirect measure. However, these results also illustrate the impact on an interested readership. Similar to citations of scientific articles or patents, once a publication has been deemed important or canonised by relevant industry gatekeepers, it is more likely to endure.

These findings are also important in supporting the plausibility of our main effects on publication quantity. In particular, it addresses within-author endogeneity, as authors arguably have less choice over the timing of their peak quality than when to start their career. If the observed productivity gains were driven by the fact that moving to New York City purely coincides with the decision to start a career, it is unlikely that we would also observe these patterns of quality gains. Therefore, we can more confidently attribute these productivity and quality gains to agglomeration economies in New York City.

5 Conclusion

This paper shows how New York City offered significant productivity benefits to authors residing there during the early stages of the industry's professionalisation path. However, we also find how individual benefits and industry concentration in New York might have declined with industry growth and maturity. Given this strong growth of the publishing industry all over the United States and widely decreased costs of publication after World War II, we attribute the decreasing publication quantity differential between New York and the rest of the country primarily to a catch-up by the periphery, the professionalisation of the publishing industry, and technological progress, rather than an absolute decline of literary circles and the publishing industry in New York City. This might have also allowed the more productive and established authors in particular to move away without a significant productivity cost, lowering the output differential between the city and the rest of the country.

With our data, we cannot give a final answer to the persistence of quality rather than quantity effects. We have shown that residence in New York did not just make it more likely to publish but to publish critically acclaimed works. From a theoretical perspective, it is probable that success with peers including literary critics and editors is largely helped by social or cultural and in turn geographic proximity. Lower publication costs would not directly remedy disadvantages through social and cultural distance. Therefore, the impact on quality should be more persistent than the impact on quantity. Although, this might change the more social and professional interactions move online.

Due to the composition of our sample, our estimate cannot be seen as the average treatment effect for an average American author but the effect on a talented and relatively young subgroup of the population. However, our primary research interest is in observing the evolution of agglomeration economies in a creative sector over a long time horizon. We focus less on the composition of the beneficiaries in this paper, but we acknowledge that literary agglomeration may have been particularly beneficial to specific demographics of authors.

Furthermore, we have highlighted the importance of location in a developing industry. Naturally, the strict geographic bounds of these agglomeration effects may lead to specific ethnic and socioeconomic groups being over-represented. For instance, our estimates would imply that a

random sample of eminent authors is more likely to be representative of New York than of the overall US population, or be skewed towards historically mobile groups such as young men. In addition, social and cultural barriers as described often for literary circles would further limit diversity among successful authors. More research into the interaction of agglomeration economies and exclusionary practices could provide important insights.

References

- Arnold, Heinz Ludwig (2009). Preface to the third, completely revised edition. In H.L. Arnold (Ed.), *Kindlers Literatur Lexikon*. J.B. Metzler. <https://www.derkindler.de/das-lexikon/vorwort-kindler-2009/>.
- Arzaghi, M. and Henderson, J. V. (2008). Networking off Madison Avenue. *The Review of Economic Studies*, 75(4):1011–1038.
- Baum-Snow, N. and Ferreira, F. (2015). Causal inference in urban and regional economics. In *Handbook of Regional and Urban Economics*, volume 5, pages 3–68. Elsevier.
- Becker, G. S. (1964). Human capital: A theoretical and empirical analysis, with special reference to education. *University of Illinois at Urbana-Champaign’s Academy for Entrepreneurial Leadership Historical Research Reference in Entrepreneurship*.
- Becker, H. S. (1974). Art as collective action. *American Sociological Review*, pages 767–776.
- 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. (2015). Agglomeration economies in classical music. *Papers in Regional Science*, 94(3):443–468.
- Borowiecki, K. J. and Dahl, C. M. (2021). What makes an artist? The evolution and clustering of creative activity in the US since 1850. *Regional Science and Urban Economics*, 86:103614.
- Bourdieu, P. (1969). Intellectual field and creative project. *Social Science Information*, 8(2):89–119.
- Bourdieu, P. (1985). The market of symbolic goods. *Poetics*, 14(1-2):13–44.
- Bourrier, K. and Thelwall, M. (2020). The Social Lives of Books: Reading Victorian Literature on Goodreads. *Journal of Cultural Analytics*, 1(1).
- Buenstorf, G. and Klepper, S. (2009). Heritage and agglomeration: The Akron tyre cluster revisited. *The Economic Journal*, 119(537):705–733.
- Butts, K. (2021). *did2s: Two-Stage Difference-in-Differences Following Gardner (2021)*.

- Cabello, M. and Rojas, C. (2016). Researchers and the wealth of nations. Mpra paper, University Library of Munich, Germany.
- Cabral, L., Wang, Z., and Xu, D. Y. (2018). Competitors, complementors, parents and places: Explaining regional agglomeration in the US auto industry. Review of Economic Dynamics, 30:1–29.
- Callaway, B. and Sant’Anna, P. H. (2021). Difference-in-differences with multiple time periods. Journal of Econometrics, 225(2):200–230. Themed Issue: Treatment Effect 1.
- Cameron, A. C. and Trivedi, P. K. (2001). Essentials of count data regression. A companion to theoretical econometrics, 331.
- Carlino, G. and Kerr, W. R. (2015). Agglomeration and innovation. In Handbook of Regional and Urban Economics, volume 5, pages 349–404. Elsevier.
- Ceulemans, C., Ginsburgh, V., Prieto-Rodriguez, J., and Weyers, S. (2020). One hundred years of solitude: Bestsellers in the United States, 1900-1999. Poetics, 79:101422.
- Ciccone, A. and Hall, R. E. (1996). Productivity and the density of economic activity. American Economic Review, pages 54–70.
- Combes, P.-P. and Gobillon, L. (2015). The empirics of agglomeration economies. In Handbook of Regional and Urban Economics, volume 5, pages 247–348. Elsevier.
- de la Roca, J. and Puga, D. (2017). Learning by working in big cities. The Review of Economic Studies, 84(1):106–142.
- Diodato, D., Morrison, A., and Petralia, S. (2021). Migration and invention in the Age of Mass Migration. Journal of Economic Geography, 22(2):477–498.
- Duranton, G. and Puga, D. (2004). Micro-foundations of urban agglomeration economies. In Handbook of Regional and Urban Economics, volume 4, pages 2063–2117. Elsevier.
- Encyclopaedia Britannica (2016). Encyclopaedia Britannica. Encyclopædia Britannica Inc. www.britannica.com.
- Farrell, M. P. (2003). Collaborative circles: Friendship dynamics and creative work. University of Chicago Press.

- Gardner, J. (2022). Two-stage differences in differences.
- Giorcelli, M. and Moser, P. (2020). Copyrights and creativity: Evidence from Italian opera in the Napoleonic Age. Journal of Political Economy, 128(11):4163–4210.
- Glaeser, E. L. (2005). Urban Colossus: Why is New York America’s largest city? Economic Policy Review, page 7.
- Goodreads.com (2020). <https://www.goodreads.com>.
- Green, J. N. (2010). The book trades in the new nation. In Gross, R. A. and Kelley, M., editors, A history of the book in America: Volume 2: An extensive republic: Print, culture, and society in the new pages 75–127. University of North Carolina Press.
- Haller, P. and Heuermann, D. F. (2020). Opportunities and competition in thick labor markets: Evidence from plant closures. Journal of Regional Science, 60(2):273–295.
- Hanlon, W. W. and Heblich, S. (2021). History and urban economics. Regional Science and Urban Economics, page 103751.
- Hanlon, W. W. and Miscio, A. (2017). Agglomeration: A long-run panel data approach. Journal of Urban Economics, 99:1–14.
- Hellmanzik, C. (2010). Location matters: Estimating cluster premiums for prominent modern artists. European Economic Review, 54(2):199–218.
- Helsley, R. W. and Strange, W. C. (1990). Matching and agglomeration economies in a system of cities. Regional Science and Urban Economics, 20(2):189–212.
- Helsley, R. W. and Strange, W. C. (2002). Innovation and input sharing. Journal of Urban Economics, 51(1):25–45.
- Jaszi, P. and Woodmansee, M. (2009). Copyright in transition. In Kaestle, C. F. and Radway, J. A., editors, A history of the book in America, Volume 4: Print in motion: The expansion of printing and reading in the pages 90–101. University of North Carolina Press.
- Kaestle, C. F. and Radway, J. A. (2009). A framework for the history of publishing and reading in the United States, 1880-1940. In Kaestle, C. F. and Radway, J. A., editors,

- A history of the book in America, Volume 4: Print in motion: The expansion of printing and reading in the pages 7–21. University of North Carolina Press.
- Kindlers Literatur Lexikon (2019). Kindlers Literatur Lexikon. J.B. Metzler. <http://www.kll-online.de>.
- Klein, A. and Crafts, N. (2020). Agglomeration externalities and productivity growth: Us cities, 1880–1930. The Economic History Review, 73(1):209–232.
- Li, X., MacGarvie, M., and Moser, P. (2018). Dead poets’ property—how does copyright influence price? The RAND Journal of Economics, 49(1):181–205.
- Literature Online (2016). Literature Online. ProQuest LLC. www.lion.chadwyck.co.uk.
- Luey, B. (2009). The organization of the book publishing industry. In Nord, D. P., Rubin, J. S., and Schudson, M., editors, A history of the book in America, Volume 5: The Enduring Book: Print Culture in Postwar America, pages 29–54. University of North Carolina Press.
- Mitchell, S. (2019). London calling? Agglomeration economies in literature since 1700. Journal of Urban Economics, 112:16–32.
- Moser, P. (2012). Innovation without patents: Evidence from world’s fairs. The Journal of Law and Economics, 55(1):43–74.
- Moser, P., Voena, A., and Waldinger, F. (2014). German jewish Émigrés and us invention. American Economic Review, 104(10):3222–55.
- New York Times (1901). A young author’s romance of love; “Springtime and Harvest,” the Story of Upton Sinclair’s Life. He Wrote It and Then Eloped, and Now Wishes to Found a Library for the People. The New York Times Archive.
- O’Hagan, J. (2021). Prominent US authors, born 1800 to 1950. Poetics, page 101589.
- Puga, D. (2010). The magnitude and causes of agglomeration economies. Journal of Regional Science, 50(1):203–219.
- Roth, A. E. and Sotomayor, M. (1992). Two-sided matching. Handbook of game theory with economic applications, 1:485–541.

- Ruggles, S., Flood, S., Goeken, R., Grover, J., Meyer, E., Pacas, J., and Sobek, M. (2020). IPUMS USA: Version 10.0 [USA Full Count 1850-1940 and USA Samples 1850-2000]. IPUMS: Minneapolis, MN. <https://doi.org/10.18128/D010.V10.0>.
- Saxenian, A. (1996). Inside-out: Regional networks and industrial adaptation in Silicon Valley and Route 128. *Cityscape*, pages 41–60.
- Schich, M., Song, C., Ahn, Y.-Y., Mirsky, A., Martino, M., Barabási, A.-L., and Helbing, D. (2014). A network framework of cultural history. *science*, 345(6196):558–562.
- Serafinelli, M. and Tabellini, G. (2022). Creativity over time and space. *Journal of Economic Growth*, 27(1):1–43.
- Stark, O. and Bloom, D. E. (1985). The new economics of labor migration. *American Economic Review*, 75(2):173–178.
- Starrett, D. (1978). Market allocations of location choice in a model with free mobility. *Journal of Economic Theory*, 17(1):21–37.
- Tao, J., Ho, C.-Y., Luo, S., and Sheng, Y. (2019). Agglomeration economies in creative industries. *Regional Science and Urban Economics*, 77:141–154.
- The Literary Encyclopedia (2016). The Literary Encyclopedia. The Literary Dictionary Company Ltd. www.litencyc.com.
- US Census Bureau (2020). Census instructions. https://www.census.gov/history/www/through_the_decades/census_instructions/.
- Waldinger, F. (2012). Peer effects in science: Evidence from the dismissal of scientists in nazi germany. *Review of Economic Studies*, 79(2):838–861.
- West III, J. L. (1989). *American authors and the literary marketplace since 1900*. University of Pennsylvania Press.
- Wikipedia Contributors (2019). Publishers Weekly lists of bestselling novels in the United States. Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/w/index.php?title=Publishers_Weekly_lists_of_bestselling_novels_in_the_United_States&oldid=986386855. Retrieved June 2019.

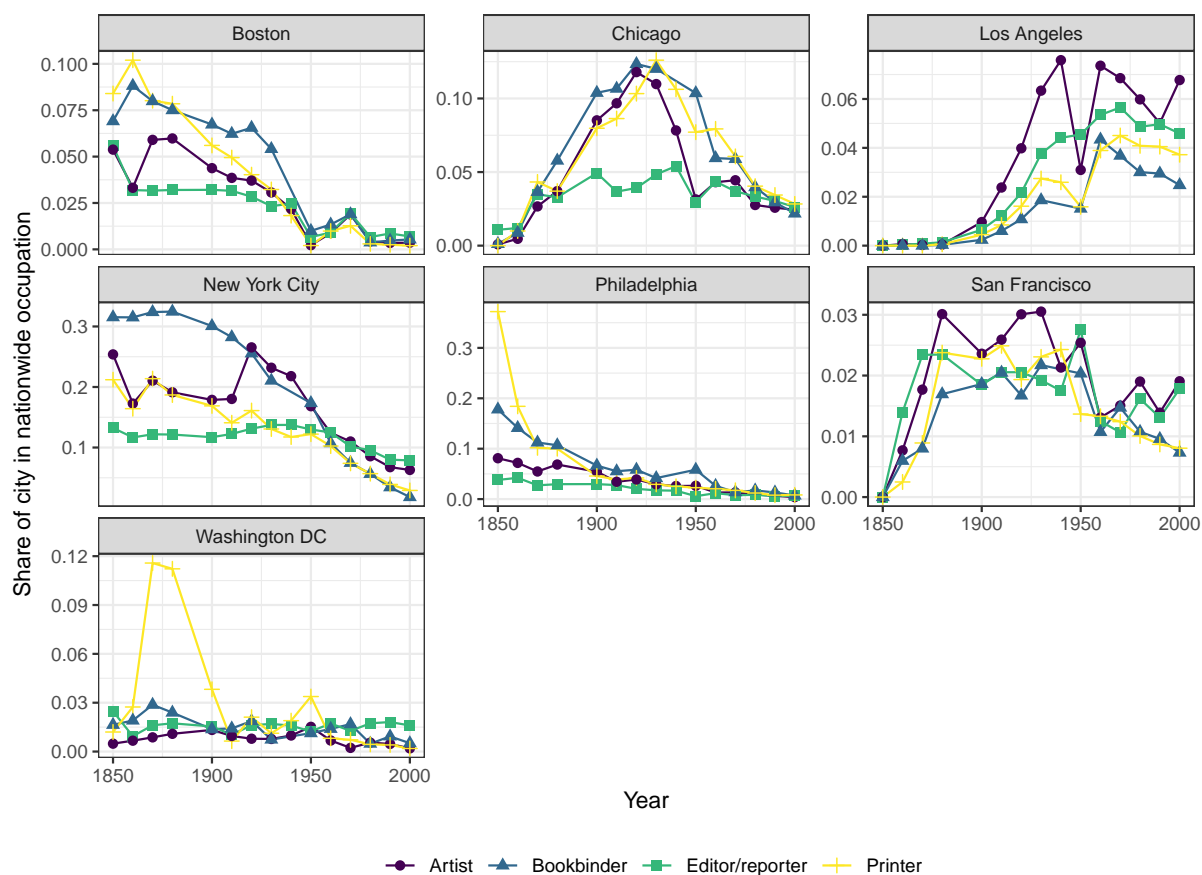
Williams, S. S. (2007). Authors and literary authorship. In Casper, S. E. and et al., editors, A history of the book in America, Volume 3: The industrial book, 1840-1880, pages 90–116. University of North Carolina Press.

Winship, M. (2009). The rise of a national book trade system in the United States. In Kaestle, C. F. and Radway, J. A., editors, A history of the book in America, Volume 4: Print in motion: The expansion of printing and reading in the pages 56–77. University of North Carolina Press.

Åke E. Andersson, Andersson, D. E., Daghbashyan, Z., and Hårsman, B. (2014). Location and spatial clustering of artists. Regional Science and Urban Economics, 47:128–137. SI: Tribute to John Quigley.

A Figures

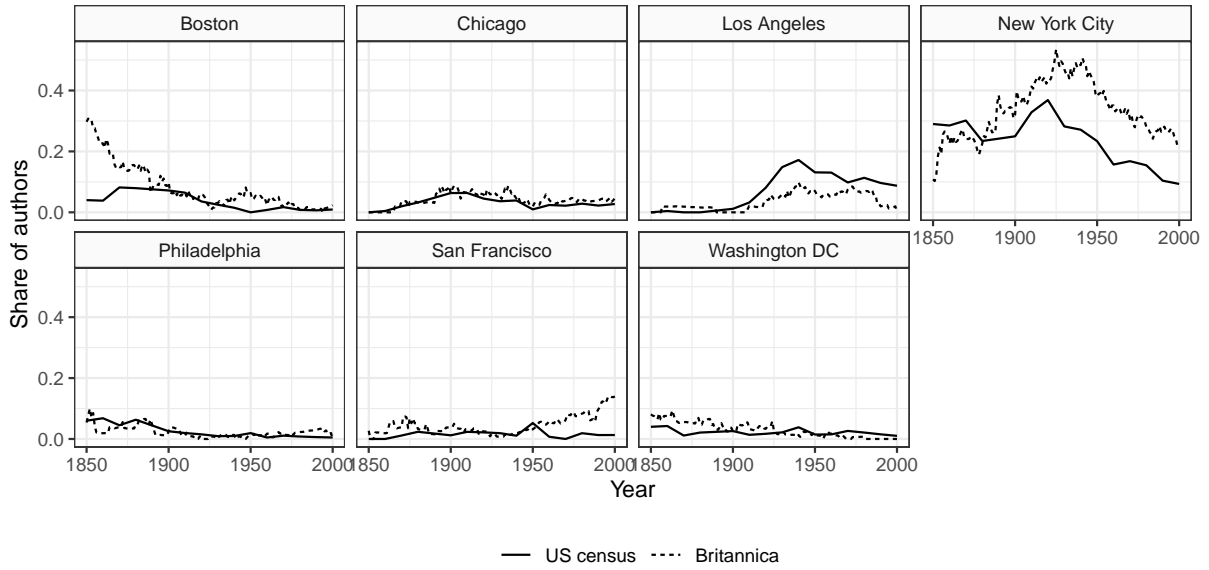
Figure 1: Major cities for publishing-related occupations



Notes: The subfigures show shares of cities in the national employment of four occupation groups: visual artists, bookbinders, editors/reporters, and printers following the IPUMS OCC1950 classifications as described in Section 3. The gap between the sum of the shares shown and 1 is the share of other locations.

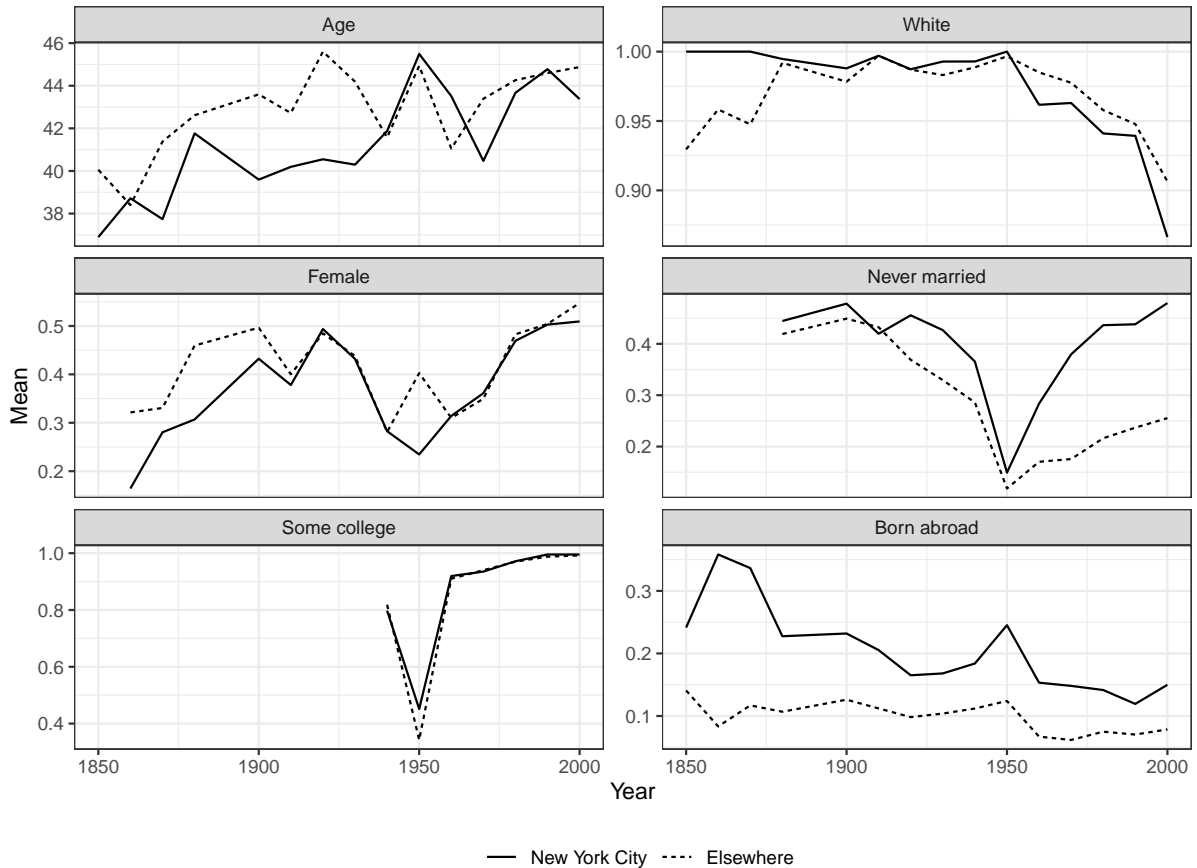
Source: Own calculation of employment shares by city based on Ruggles et al. [2020] US census data (IPUMS) data.

Figure 2: Share of authors living in major cities



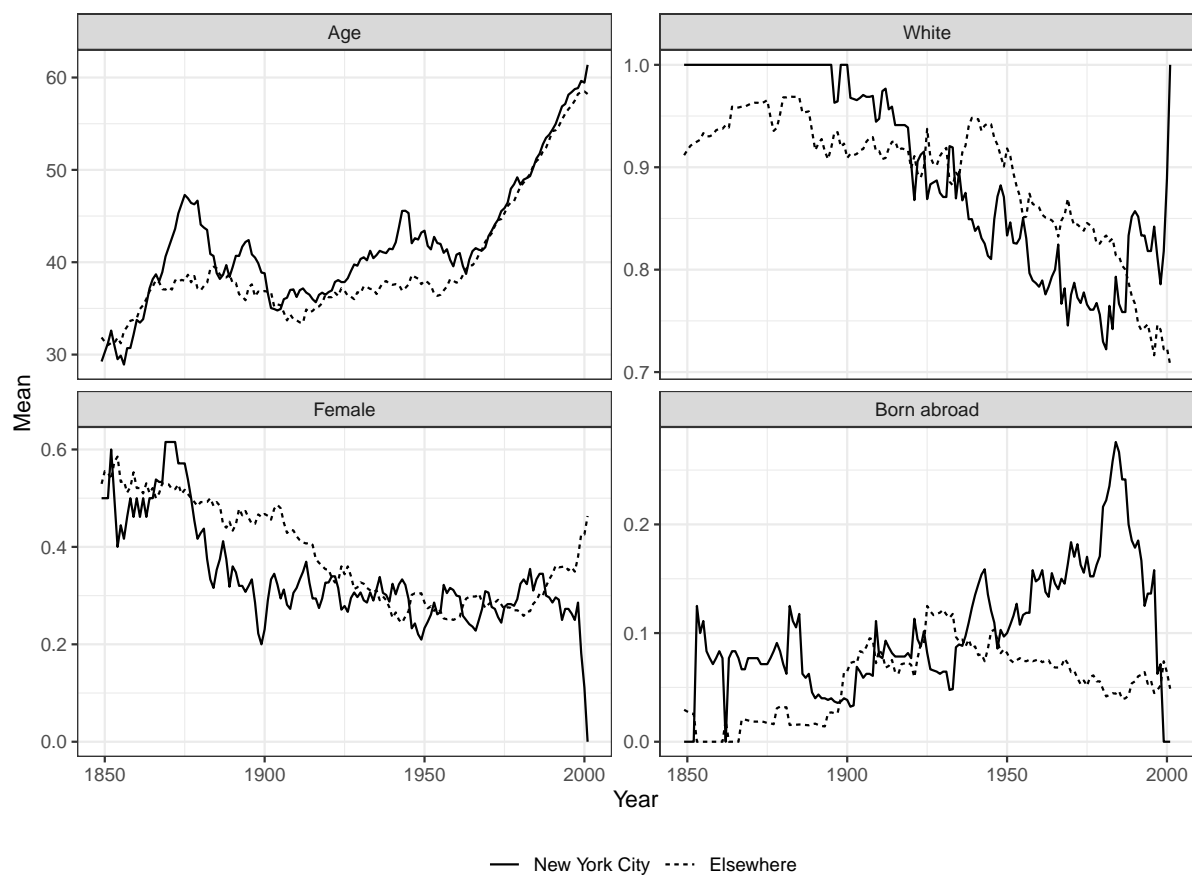
Notes: The subfigures show shares of cities among authors either listed in the US Census or authors listed in the Encyclopaedia Britannica. The gap between the sum of the shares shown and 1 is the share of other locations.
Source: Own calculation of shares of total US authors in source based on Ruggles et al. [2020] US census data (IPUMS) and *Britannica* data.

Figure 3: Characteristics of *Census* authors in New York City and elsewhere



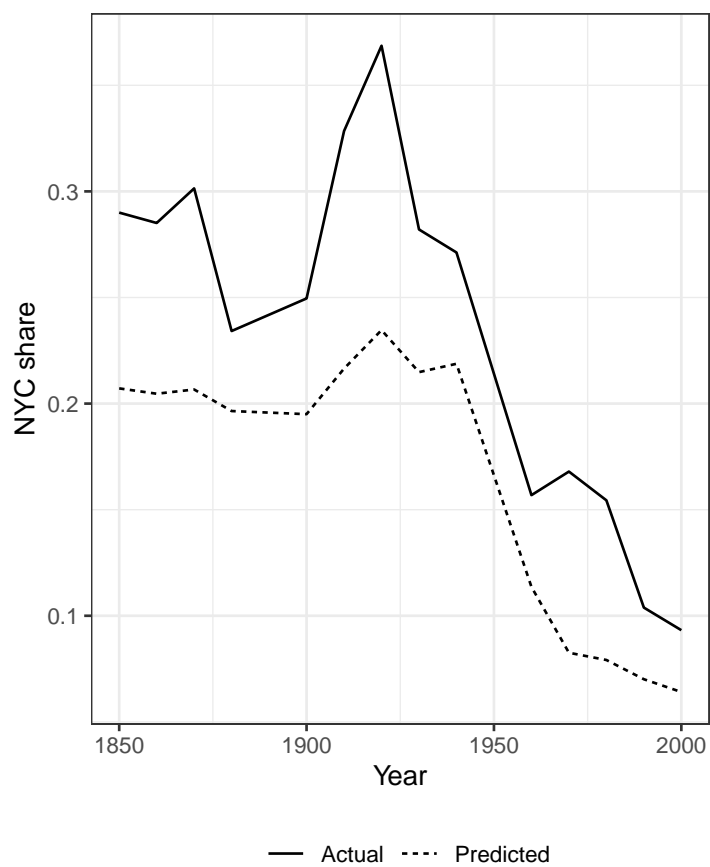
Notes: The figures show the mean age and the means of five dummy variables as given by Ruggles et al. [2020]. The calculations are based on the following variables in Ruggles et al. [2020]: AGE, RACE, SEX, MARST, EDUC, BPL.
Source: Own calculation based on Ruggles et al. [2020] US census data (IPUMS).

Figure 4: Characteristics of *Britannica* authors in New York City and elsewhere



Notes: The figures show the mean age and of dummy variables for being white, female, and born abroad over time.
 Source: Own calculation variables based on *Britannica* data.

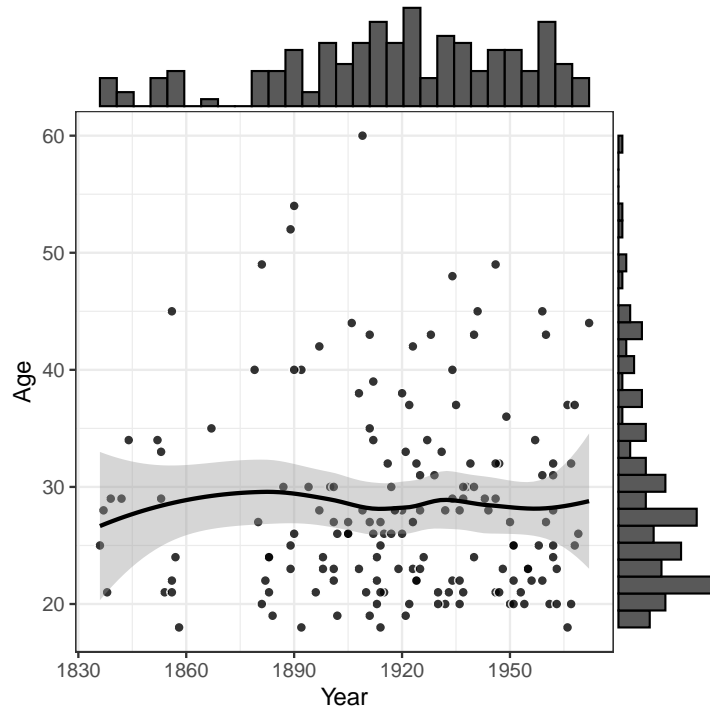
Figure 5: Author concentration in New York City



Notes: The figure shows the share of authors living in New York City following the IPUMS OCC1950 classifications as described in Section 3. The estimated share is based on the link between socioeconomic factors and author numbers estimated using counties outside NYC.

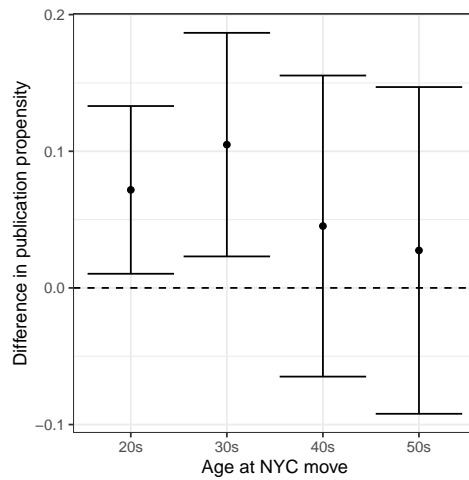
Source: Own calculations based on Ruggles et al. [2020] US census data (IPUMS) as described in Section 3.3.

Figure 6: Distribution of age and year for working age moves to New York City



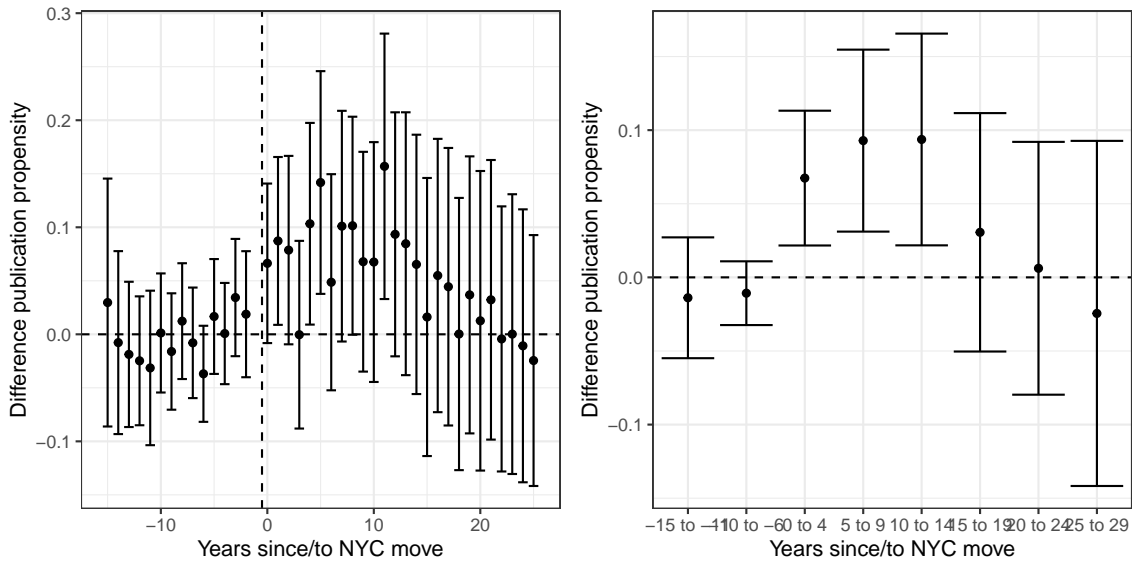
Notes: The plot shows the year and age of an author’s first move to NYC. Moves are restricted to working age: 18-65. The trend is estimated using local regressions (LOESS with a span of 0.75).
 Source: Own calculations based on *Britannica* data.

Figure 7: New York City effect by age at move



Notes: The NYC publication bonus by age of move estimated using Gardner [2022]-type regressions as described in Section 4.1. The bars represent 95% confidence intervals.

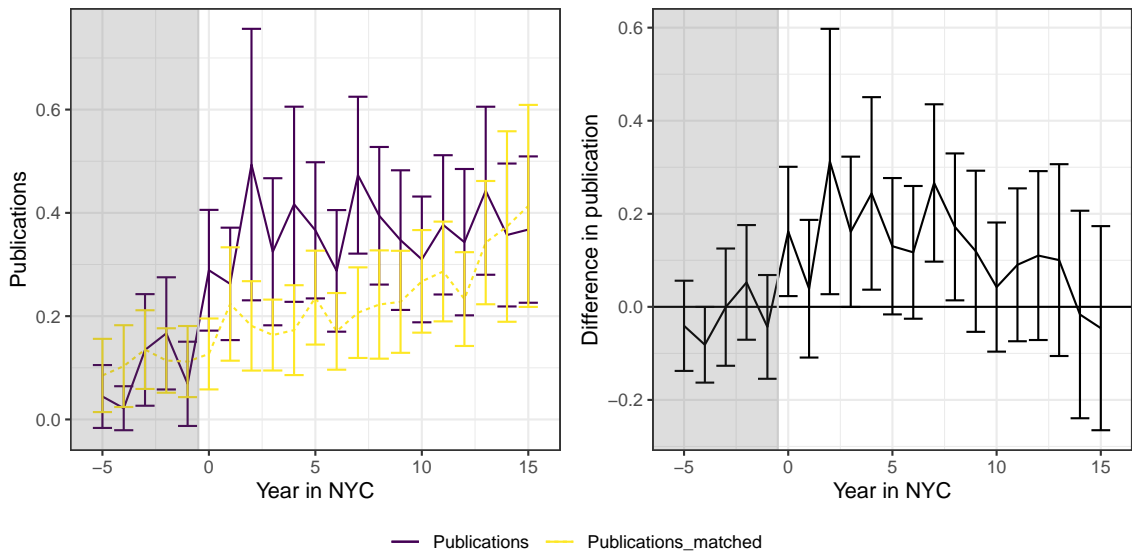
Figure 8: Productivity effect for NYC-movers



Notes: The Figure shows estimated productivity effects for writers moving to NYC relative to the move. The left plot shows individual years, the right plot five-year intervals. The reference period is the year before the move (left plot) or the five years before the move (right plot).

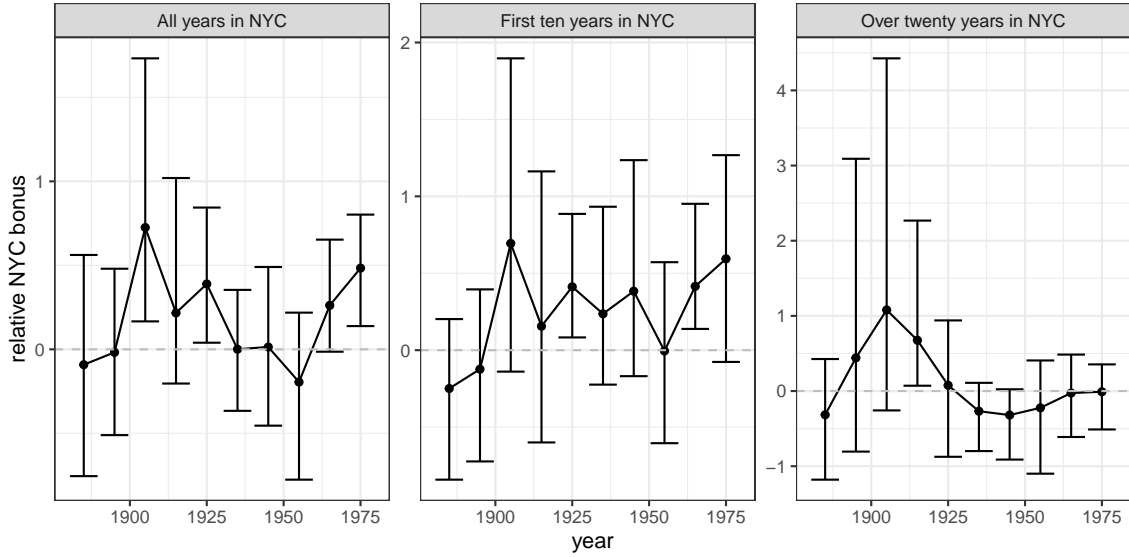
Source: Own calculations based on Gardner [2022].

Figure 9: Productivity of NYC-movers and matched authors elsewhere



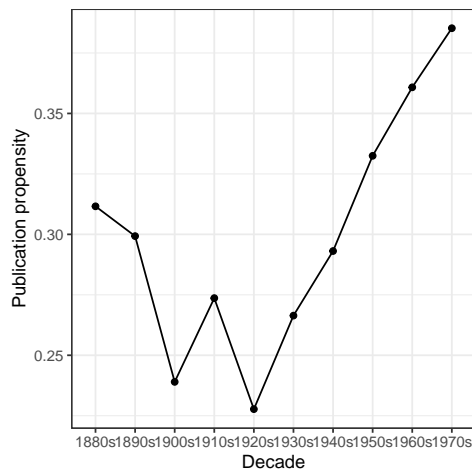
Notes: Own calculations based on coarsened exact matching as described in Section 4.1. Authors are matched on age, year (± 1 years), ethnicity, and gender. The graph shows the sample means of publications for NYC-movers and matched authors elsewhere. The bars represent 95% confidence intervals.

Figure 10: New York City effect by decade



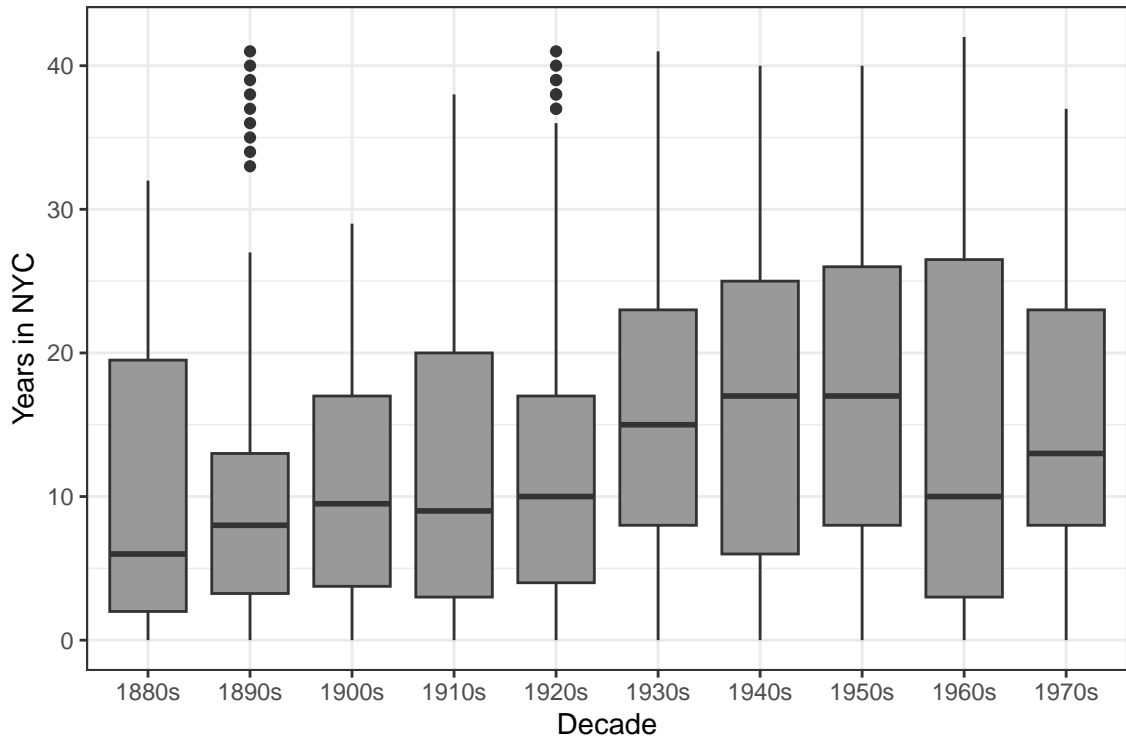
Notes: The NYC differential by decade estimated using Gardner [2022]-type regressions as described in Section 4.1. The estimated differential is divided by the decade’s average productivity to adjust for the rising publication propensity. The bars represent 95% confidence intervals from a cluster bootstrap (500 repetitions, sampled at author level).

Figure 11: Publication propensity outside New York City by decade



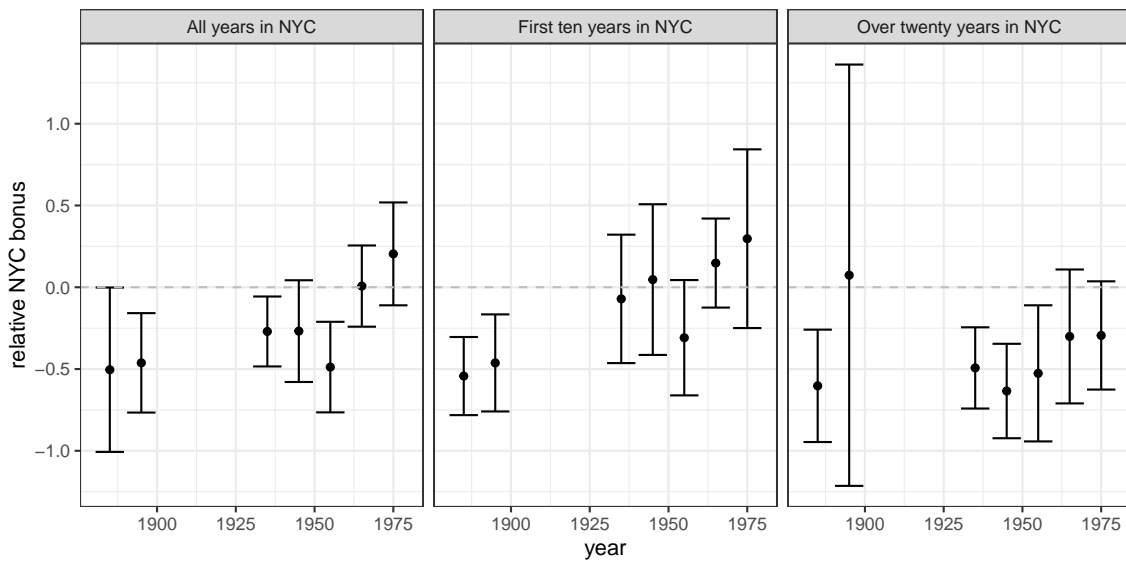
Notes: Share of publishing writers in the US outside NYC aged between 30 and 45.
Source: Own calculations.

Figure 12: Years since moving to New York City



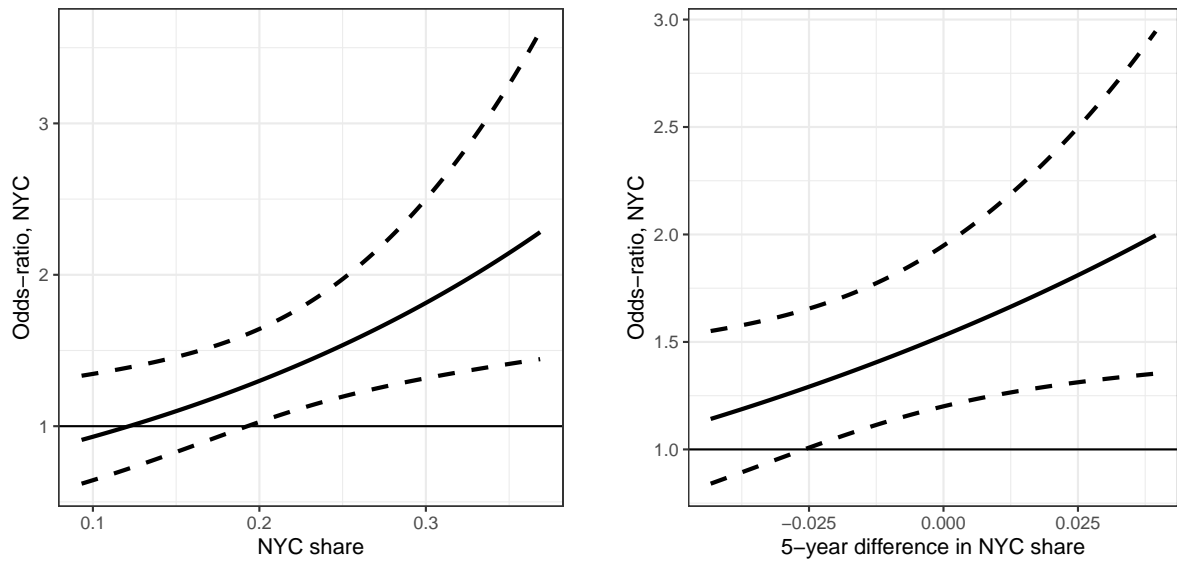
Notes: Years in NYC: years since move for writers who moved to NYC aged 23-61.
 Source: Own calculations based on Britannica sample.

Figure 13: Decade effect relative to 1900-1929



Notes: The plots show differences between the decade-specific NYC effect and a 30 year base period (1900 to 1929), estimated using Gardner [2022]-type regressions as described in Section 4.1. The estimated differential is divided by the decade's average productivity to adjust for the rising publication propensity. The bars represent 95% confidence intervals.

Figure 14: New York City effect by author concentration



(a) Relative effect by absolute author concentration (b) Relative effect by recent change in author concentration

Notes: Relative effect of living in New York City on publication propensity by yearly share of authors living in New York City and by recent change in author concentration in New York City.

B Tables

Table 1: Summary Statistics for author datasets

	Britannica	Goodreads	Kindler	Publishers Weekly	Census
Authors	471	352	171	79	47619
NYC authors	269	206	102	53	8645
% years in NYC	56	55	55	57	n/a
Female authors	159	106	47	28	n/a
Years covered	1820 - 2014	1821 - 2014	1820 - 2013	1852 - 2014	1850 - 2000
Publications	4258	1388	226	164	n/a
Publication years	1826 - 2013	1827 - 2013	1824 - 1998	1895 - 2011	n/a

Notes: All statistics refer to yearly observations of authors with a known location in the US and aged 18 to 65. Source: Own calculations based on data collection process described in Section 3.

Table 2: Probability differential of publishing while living in New York City

Dependent Variable:	At least one publication in year					
	(1)	(2)	(3)	(4)	(5)	(6)
	<u>All years in NYC</u>					
NYC	0.049*	0.046	0.062***	0.444**	0.049*	0.044
	(0.021)	(0.034)	(0.017)	(0.137)	(0.021)	(0.023)
Obs	11416	11442	11442	11170	11416	11416
	<u>First 10 years in NYC</u>					
NYC	0.078***	0.096**	0.078***	0.542***	0.076**	0.066**
	(0.023)	(0.035)	(0.018)	(0.15)	(0.024)	(0.022)
Obs	10111	10111	10111	9703	10111	10111
	<u>Over 20 years in NYC</u>					
NYC	-0.005	0.007	-0.01	0.059	-0.002	-0.001
	(0.033)	(0.041)	(0.027)	(0.269)	(0.034)	(0.03)
Obs	9866	9883	9883	9387	9866	9866
Method	Gardner	Call-SA	OLS	Logit	Gardner	Gardner
Additional controls	No	No	No	No	No	Yes
Time FE	Decade	No	Decade	Decade	Year	Year
Age & Author FE	Yes	Yes	Yes	Yes	Yes	Yes

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; $p < 0.1$

Notes: The columns show estimates using Gardner [2022], Callaway and Sant'Anna [2021], OLS, and Logit regressions with a binary dependent variable indicating a publication by an author in a year. Standard errors are clustered on the author level and reported in parentheses. Standard errors for Gardner [2022] and Callaway and Sant'Anna [2021] estimations are bootstrapped, otherwise analytical. Additional controls are a dummy for a move in the last five years, the sum of moves up to the current year, and a dummy for having published previously.

Table 3: Other large cities and moves away from NYC

Dependent Variable:	At least one publication in year			
	(1)	(2)	(3)	(4)
Model:				
Other Large City	0.0232	0.0367		
	(0.0269)	(0.0295)		
NYC			0.0490**	
			(0.0222)	
After NYC			0.0365	
			(0.0277)	
NYC or after NYC				0.0445**
				(0.0184)
Observations	7,117	6,635	12,674	12,674
Max years since move	Any	10	Any	Any
Time FE	Decade	Decade	Decade	Decade
Age & Author FE	Yes	Yes	Yes	Yes

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; $p < 0.1$

Notes: The columns show estimates using Gardner [2022] type regressions with a binary dependent variable indicating a publication by an author in a year. Standard errors are clustered on the author level and reported in parentheses.

Table 4: Critical, commercial, and modern readership success

Dependent Variable:	Kindler (1)	Publishers Weekly (2)	> median (3)	ln(reviews) (4)
NYC	0.0345** (0.0150)	0.0342 (0.0408)	0.2066*** (0.0765)	1.296*** (0.4924)
Observations	3,739	1,740	705	705
$E(Y \mid \text{outside NYC})$	0.037	0.067	0.544	33,237

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; $p < 0.1$

Notes: The columns show estimates using Gardner [2022] type regressions with a dummy indicating a Kindler or Publishers Weekly publication by an author in a year (Columns (1) and (2)), a dummy indicating more than the median number of reviews on Goodreads in column (3), or the logarithm of the number of reviews on Goodreads. The unit of observation differs between the columns: years of authors with at least one Kindler publication (Column 1), years of authors with at least one Publishers Weekly publication (Column 2), novels listed on Goodreads (Columns 3 and 4). Standard errors are clustered on the author level and reported in parentheses.