DISCUSSION PAPER SERIES

DP14686

MODERN LIBRARY HOLDINGS AND HISTORIC CITY GROWTH

Eric Chaney

ECONOMIC HISTORY

ISSN 0265-8003

MODERN LIBRARY HOLDINGS AND HISTORIC CITY GROWTH

Eric Chaney

Discussion Paper DP14686
Published 30 April 2020
Submitted 29 April 2020

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

MODERN LIBRARY HOLDINGS AND HISTORIC CITY GROWTH

Abstract

This paper uses more than 30 million records from the union of the world's largest library collections to provide a novel proxy for historic economic activity. Changes in the number of authors affiliated with a city strongly correlate with existing population growth estimates. An empirical Bayes approach exploits this correlation to improve upon current data and provide new growth estimates where none exist. The paper concludes by using the new data to revisit urban growth during the rise of the Atlantic trade.

JEL Classification: N/A

Keywords: N/A

Eric Chaney - eric.chaney@economics.ox.ac.uk
University of Oxford and CEPR

Acknowledgements
This project was inspired by a seminar comment by RobertJ.Barro. I thank Sascha Becker, David de la Croix, James Fenske, Jesús Fernández-Villaverde, Joel Mokyr, Martin Weidner and participants at Cambridge, Groningen and OWL for helpful discussions and comments. Part of this research was carried out while I was Furer fellow at Harvard University, whose financial support is gratefully acknowledged. All remaining errors are mine.

Estimates of historical city sizes underpin our understanding of long-run economic development. For example, changes in urbanization rates are often used as a proxy for the evolution of economic prosperity in the past (e.g. De Long and Shleifer, 1993; Acemoglu et al., 2002, 2005; Dittmar, 2011; Nunn and Qian, 2011; Bosker et al., 2013). Despite the widespread use of this proxy, we lack city size estimates for many historical regions. Where estimates are available they are often no more than educated guesses, subject to errors of 30 percent or more (e.g. Bairoch et al., 1988, pp. ix, 298).

This paper uses modern library holdings to provide a new proxy for historical city growth. Panel A of Figure 1 illustrates this metric by plotting the evolution of existing London population estimates and the number of London-based authors over time. The co-movement of the two series is consistent with historical (e.g. Mokyr, 1993) and empirical evidence (e.g. Glaeser et al., 2004; Dittmar, 2011) linking city growth and intellectual production. It also suggests that author growth can be used as an additional measure of city growth.

Although author growth is a noisy proxy for city growth, Henderson et al. (2012) demonstrate how such additional proxies can be used to improve existing city growth estimates. Building on this insight, I derive the weighting proposed by Henderson et al. (2012) within an empirical Bayes framework. I then use this framework to generate both new and improved estimates of city growth. Panel B graphs these estimates for London, demonstrating how this approach generates growth estimates where none currently exist. I illustrate the new data's efficiency gains by revisiting Acemoglu et al. (2005)'s influential study. The paper then highlights the high-frequency nature of the data by tracking the Atlantic traders' divergence from the rest of Western Europe by year. This exercise provides preliminary evidence that this process began in the early part of the sixteenth century.

To my knowledge, this is the first paper to use author counts to proxy for city growth. Nevertheless, in recent years researchers have exploited large databases of notable individuals and authors to measure both the evolution of longevity and migration networks over time (Schich et al., 2014; de la Croix and Licandro, 2015; Cummins, 2017; de la Croix et al., 2019). Within this literature, Gergaud et al. (2017) is most closely related to this paper, using information on a little over one million notable people to explore correlations between city growth and a variety of characteristics of notable individuals. This study, however, does not find a robust association between notable individuals and city growth nor does it use a statistical framework to improve upon existing growth data.

More broadly, the paper's central premise is consistent with work relating book production to both urbanization and economic development (e.g. Baten and van Zanden, 2008; Buringh and van Zanden, 2009). Distinct from these studies, I suggest how author counts can be used to improve upon existing city growth estimates, thus contributing to the literature using modern econometric methods to enhance our understanding of urban patterns in the distant past (e.g. Barjamovic et al., 2019).

From a conceptual standpoint this paper can be viewed as an extension of the rapidly growing luminosity literature back in time (for an overview of this literature see Michalopoulos and Papaioannou, 2018). It is novel in that it links this literature to a Bayesian framework (e.g. Kane and Staiger, 2008; Chetty et al., 2014; Angrist et al., 2017; Fessler and Kasy, 2019), exploiting advances in online data availability and computation over the past decades (e.g. Mullainathan and Spiess, 2017) to provide a new proxy for historic economic activity.

The remainder of the paper proceeds as follows. I begin by describing the data and providing examples of how they can be used to measure variation in economic activity. A second section introduces the conceptual framework, deriving the Henderson et al. (2012) weighting within an empirical Bayes framework. The third section details the empirical analysis and revisits Acemoglu et al. (2005). A final section concludes.

1 Library Data

Author counts derive from the "authority clusters" contained in the Virtual International Authority File (VIAF) database (www.viaf.org).[1] These clusters are groupings of authority files. Authority files, in turn, are created by libraries to assign a uniform name and unique identifier to a given author's works and generally contain (among other entries) a numerical identifier, an author's authoritative name and name variants (Hickey and Toves, 2014). For example, in the Library of Congress the name **Avicenna, 980-1037** is the authoritative name for the medieval philosopher. Although individual works by this author may bear variants on this name (such as **Abu Ali ibn Sino, 980-1037**) they are catalogued under the authoritative name **Avicenna, 980-1037**.

VIAF was created, in part, to provide authority clusters (or uniform authority files) across participating libraries. To better understand these clusters consider the Iranian astronomer Shams al-Din

[1] This section provides a brief description of the data. See the online appendix for a more detailed treatment.

al-Khafri. This author's works are held in the Library of Congress under the authoritative name **Khafrī, Muḥammad ibn Aḥmad, active 1526** and in the National Library of the Netherlands under **Ḵafrī, Šams al-Dīn Muḥammad b. Aḥmad**. VIAF's authority cluster numbered 120564917 links these names and provides author information contained in the constituent libraries' authority files.

This database began as a joint project between the Library of Congress, the Deutsche Nationalbibliothek, the Bibliothèque Nationale de France and the Online Computer Library Center. Many of the world's largest libraries have since joined this initial group and VIAF is regularly updated as the number of participating libraries and their collections change. As of April 2014 VIAF contained 26 million authority clusters built from 38 million authority records along with 104 million bibliographic records associated with these authority records (Hickey and Toves, 2014). In this paper I use the 30,543,006 authority clusters contained in the VIAF database as of the 7th of August 2017.

Starting from this collection of clusters, I constructed the baseline data as follows. First, I removed records that did not contain at least one numerical character in the birth or death fields as well as those containing birth or death years after 1799. I then checked the 594,741 remaining clusters for errors against name variants in the authority files and duplicate clusters were removed as were authors from outside the Europe and the Middle East and North Africa (MENA) regions.[2] When this was done 461,612 clusters with death dates on the interval [800,1800) remained.

To geo-reference authors, I relied on seven geo-referencing sources and a mix of automated and manual processes. The geo-referencing sources can be broadly grouped into information contained in the catalogues of participating libraries, Wikipedia and the authority clusters themselves.[3] Using these sources I was able to geo-reference 243,451 of the remaining 461,612 clusters.

Panel A of Table 1 gives a sense of the time dimension of the geo-referenced data, detailing author counts by century for 9 select cities. The general increase in author counts over time is apparent, although not uniform. For example, in Baghdad the author count in 1700 is less than one tenth that of 900. In panel B I include existing city size estimates for comparison. A quick inspection of these data

[2]I focus on these two regions in order to maintain focus and due to my scholarly limitations (e.g. I do not read any East Asian languages). I determined whether an author belonged to the Europe/MENA region based on names: for example, authors with Islamic-sounding names were assigned to the MENA region (thus, for the purposes of this paper MENA is synonymous with the Islamic world). This name-based classification seems to be quite accurate: for the subset of geo-referenced authors I only incorrectly assigned a handful of authors (primarily those with Jewish names).

[3]To illustrate this last source, consider the authority cluster 27456126 which contains the name variant "Adalbero Virdunensis -1158" suggesting this author had a connection to the city of Verdun in modern-day France.

suggests that the relationship between author counts and city size is city-specific. For this reason, my focus in this paper is on using author growth to measure city growth.

Figure 2 depicts the distribution of geo-referenced authors across space, giving the locations of cities with at least one author death on the interval [800,1800). In this figure, larger grey circles denote more observed author deaths and black squares give the locations of VIAF libraries. The density of authors in Western Europe is striking as is the concentration of authors around Middle Eastern urban centers.

After this geo-referencing process was complete, I limited the data to the 150,906 clusters whose authors died in the 1,432 cities for which Bairoch et al. (1988) provide at least one population estimate on the interval [800,1700]. I restricted the sample in this manner to focus on the urban areas used in previous studies.[4] From this sample, I then excluded 111 cities in southeastern Europe (defined as being in modern-day Albania, Bosnia and Herzegovina, Bulgaria, Greece, Hungary, Kosovo, North Macedonia, Romania, Serbia and Slovakia) due to evidence that the relationship between author and city growth is different in these regions as discussed in the online appendix in greater detail. For the regression analysis this leaves 1,843 Bairoch et al. (1988) growth estimates.[5]

Figure 3 provides two high-frequency examples of the relationship between author and city growth. Panel A provides the evolution of the logarithm of one plus author counts in Magdeburg, Germany for years after 1500. In this panel the black line denotes smoothed deaths (using a rear moving average of 30 years). The vertical line marks the sack of the city in 1631 during which the city's population is believed to have declined by 83% (from 30,000 to approximately 5,000 inhabitants (McLeod, 2001)). Panel A is consistent with this estimate in the sense that author counts experience a sharp drop in the years following the massacre.

Panel B illustrates the evolution of author counts in Avignon, France from 1200 onwards. The first vertical line denotes the movement of the papacy from Rome to the city in 1309 which quickly transformed the "village into a vibrant cosmopolitan city" (Rollo-Koster, 2015, p. 188) of at least 30,000 inhabitants (Zutshi, 2000, p. 669).[6] The second line denotes the deposition of the Avignon pope by the Council of Pisa in 1409 (Rollo-Koster, 2015, p. 270) which was followed eight years later by the

[4]It seems plausible, however, that author counts could also be used to measure economic activity in more rural areas. I leave this possibility to future research.

[5]For the MENA regressions, I begin from the Bosker et al. (2013) population estimates and limit attention to the 424 growth estimates that can be obtained for cities which they classify as Muslim (in both relevant centuries) on the interval [800,1700].

[6]This rapid growth rendered Avignon the largest town in France after Paris (Zutshi, 2000, p. 669).

end of the Avignon papacy. The rapid jump and subsequent fall in the number of Avignon authors is consistent with these historical events, demonstrating how concentration of political power can lead to rapid urban growth (e.g. Ades and Glaeser, 1995).

How accurate is the geo-referencing process? To explore this question, I drew a random sample of 1000 authors and manually verified the geo-referencing information using the seven sources described above. I found that 83% of authors were correctly geo-referenced (95% confidence interval of [0.81,0.86]). This seems reasonable, particularly in light of the fact that many discrepancies stem from ambiguous situations where death places must be imputed from multiple places of activity.

While it is natural to question the selection process underlying the VIAF database, I claim that to a first approximation the data represent the population of all authors whose works have survived until today. This claim is rooted in the observation that many constituent VIAF libraries aspire to gather the corpus of known intellectual production in both their home countries and beyond (e.g. https://www.loc.gov/about).

Although testing this claim in a global sense is impossible, I investigated its empirical implications by drawing a random sample of 1000 print editions contained in the Universal Short Title Catalogue database of St. Andrews (2019) (https://ustc.ac.uk/).[7] This source aims to cover all books published in Europe between the invention of printing and 1650 and as such represents the population of all authors whose printed works have survived until today. I successfully located 81% of these authors (95% confidence interval of [0.79,0.83]) in the baseline data. Such a high level of coverage is consistent with the claim that VIAF provides a reasonable approximation to the population of known European authors.

Nevertheless, it is undoubtedly true that survival probabilities are higher in some regions and time periods than in others. The following section develops a simple conceptual framework to make clearer how such issues may affect the analysis.

[7] I scraped these data between third and the sixth of October 2019. Dittmar and Seabold (2019) quantitatively analyze these data.

2 Conceptual Framework

I begin by demonstrating that changes in author counts can provide information on city growth under minimal assumptions. For example, suppose that in period t one observes X^*_{jt} scholars dying in city j and that this count is related to the population Y_{jt} by the relationship $X^*_{jt} = \delta_{jt} Y_{jt}$, $\delta_{jt} > 0, Y_{jt} > 0$. These assumptions yield:

$$\Delta E[ln(X^*_{jt})] = \Delta E[ln(\delta_{jt})] + \Delta E[ln(Y_{jt})] \tag{1}$$

where Δ denotes first differences. Letting D denote treatment in a differences-in-differences setup, equation 1 implies that $sgn(\Delta E[ln(X^*_{jt})|D = 1] - \Delta E[ln(X^*_{jt})|D = 0]) = sgn(\Delta E[ln(Y_{jt})|D = 1] - \Delta E[ln(Y_{jt})|D = 0])$ as long as the way the average of the logarithm of δ_{jt} evolves in the treatment and control groups is not dissimilar. Thus, differences-in-differences estimates using author counts can be used to ascertain whether a group of cities grew faster than another in many empirical environments.

Although this discussion relies on few assumptions, it also ignores empirical regularities in the urban growth process. Incorporating such regularities into the framework yields useful insights.

It seems reasonable to assume that the number of authors C_j working in city j is one of the many urban properties which vary with population Y_j according to a scaling relation of the form $C_j = C_{0j} Y_j^{\beta_j}$ (e.g. Gomez-Lievano et al., 2012). Taking logarithms and first differences yields:

$$c_{jt} = \beta_j \tilde{y}_{jt} \tag{2}$$

where $c_{jt} = ln(C_{jt}) - ln(C_{jt-1})$, $\tilde{y}_{jt} = ln(Y_{jt}) - ln(Y_{jt-1})$ and $var(\tilde{y}_j) = \sigma^2_y$.

I do not observe c_{jt} but can measure the difference in number of authors dying in a given city whose works have survived until today $\tilde{x}^*_{jt} = ln(X^*_{jt}) - ln(X^*_{jt-1})$.[8] To understand the implications of using this proxy, assume without loss of generality that authors are homogenous and die with probability p and if an author dies his works survive to the present with probability p_t.[9] Observed author deaths X^*_{jt} will then follow a binomial distribution with parameters C_{jt} and time-varying observation probability

[8] Note the only source of measurement error in X^*_{jt} is that induced by survival. Below I will allow for other sources of measurement error.

[9] Given the high mortality rates in pre-industrial cities, assuming that the hazard is constant across all authors seems a reasonable simplifying assumption.

$\pi_t = p \cdot p_t$. It is straightforward to show (e.g. Katz et al., 1978; Koopman, 1984) that as C_{jt} grows $\tilde{x}^*_{jt} = ln(\frac{\pi_t}{\pi_{t-1}}) + c_{jt} + \tilde{u}_{jt}$ where $E[\tilde{u}_{jt}|c_{jt}] = 0$. In other words, using changes in author deaths as a proxy for the growth in the number of authors working in a given city leads to classical measurement error. This discussion also demonstrates how changes in the probability that an author's works are held in the VIAF collection can be absorbed by time fixed effects.

Clearly, this is not the only source of measurement error in author counts, as libraries often provide only approximate death dates. In addition, places of death are also measured with noise. Consequently, I assume that observed author growth $\tilde{x}_{jt} = \tilde{x}^*_{jt} + \tilde{\eta}_{xjt}$ where σ^2_{xt} denotes the variance of the error term.

The expressions for $\tilde{x}_{jt}$, $\tilde{x}^*_{jt}$ and equation 2 yield:

$$\tilde{x}_{jt} = \beta \tilde{y}_{jt} + \tilde{\epsilon}_{xjt} \tag{3}$$

where $\tilde{\epsilon}_{xjt} = ln(\frac{\pi_t}{\pi_{t-1}}) + \tilde{u}_{jt} + \tilde{\eta}_{xjt} + (\beta_j - \beta)\tilde{y}_{jt}$. Let $\epsilon_{xjt}, y_{jt}, x_{jt}$ denote the time demeaned values of each variable and write equation 3 as $x_{jt} = \beta y_{jt} + \epsilon_{xjt}$. The assumptions that $z_{jt} = y_{jt} + \epsilon_{zjt}$ (where ϵ_{zjt} is classical measurement error with variance σ^2_{zt}) and $cov(\epsilon_{zjt}, \epsilon_{xjt}) = 0$ are sufficient to apply the framework in Henderson et al. (2012).

The key assumption underpinning this framework is the lack of correlation between errors in existing city growth estimates and ϵ_{xjt} as discussed in detail below. To the extent to which this assumption holds, the conceptual discussion could end here and the optimal weighting derived in equation 4 below would be valid.

2.1 Empirical Bayes

For expositional simplicity, I begin by assuming (dropping time subscripts for now) that measured city growth conditional on true growth is distributed $N(y_j, \sigma^2_z)$ and $y_j \sim N(\psi x_j, \sigma^2_y(1-\rho^2_{xy}))$ where ρ_{xy} is the correlation between x and y and ψ is the population regression coefficient from the regression if y_j on x_j.[10] While these assumptions are at best an approximation to reality, they allows for a clear illustration of the link between Henderson et al. (2012) and the empirical Bayes approach. It is important to stress, however, that this link is robust to violations of these distributional assumptions (which will be relaxed

[10]Note that even though ψ is biased (due to measurement error in author growth), ψx_j is the best fit relationship as noted in Henderson et al. (2012). See below for a more detailed discussion.

below).

Under these assumptions ψx_j can be viewed as the mean of the prior distribution, z_j the observed data and $z_j|y_j$ the likelihood. Standard arguments (e.g. Fay III and Herriot, 1979; Morris, 1981) show that the convex combination of existing estimates and OLS fitted values $\frac{[\sigma_y^2(1-\rho_{xy}^2)]z_j+\sigma_z^2(\psi x_j)}{\sigma_z^2+\sigma_y^2(1-\rho_{xy}^2)}$ yields a more precise estimate of y_j (in a mean squared sense) than existing growth estimates or the OLS fitted values on their own.

This weighted average is numerically identical to the optimal weighting in Henderson et al. (2012). To see this, write the weight on the existing estimate of city growth as $\lambda^* = \frac{\sigma_y^2(1-\rho_{xy}^2)}{\sigma_z^2+\sigma_y^2(1-\rho_{xy}^2)}$, the signal to noise ratio $\phi = \frac{\sigma_y^2}{\sigma_y^2+\sigma_z^2}$ and note that $\rho_{xz}^2 = \phi\rho_{xy}^2$. This implies that λ^* is equal to

$$\frac{\phi - \rho_{xz}^2}{1 - \rho_{xz}^2} \tag{4}$$

which is the optimal λ from equation 11 of Henderson et al. (2012). Thus, the optimal weighting of existing and regression growth estimates $\hat{y}_j = \lambda^* z_j + (1-\lambda^*)\hat{\psi}x_j$ in Henderson et al. (2012) can be viewed as the empirical Bayes or "shrinkage" estimator (e.g. Kane and Staiger, 2008; Chetty et al., 2014; Angrist et al., 2017; Fessler and Kasy, 2019). This equivalence flows from the robustness of the empirical Bayes weighting to violations of normality (e.g. Efron and Morris, 1973).[11]

From a Bayesian perspective analysis proceeds as if the prior distribution centered around the correct conditional expectation were given, the noisy estimates observed and Bayesian updating subsequently applied to obtain the posterior distribution. The posterior distribution has mean $\frac{[\sigma_y^2(1-\rho_{xy}^2)]z_j+\sigma_z^2(\psi x_j)}{\sigma_z^2+\sigma_y^2(1-\rho_{xy}^2)}$ which is the empirical Bayes estimator. This estimator "shrinks" the noisy estimates z_j back towards the prior mean. When $\sigma_y^2(1-\rho_{xy}^2)$ is large relative to σ_z^2 the shrinkage factor is small and the posterior mean is close to z_j. When the converse is true, the posterior mean is close to the prior ψx_j. In sum, all else equal the tighter true city growth is clustered around the OLS line, the higher weight the fitted values will receive in the composite estimate.

Establishing the link between the empirical Bayes approach and Henderson et al. (2012) is useful for at least three reasons. First, it allows for the derivation of Henderson et al. (2012)'s optimal weighting

[11] Henderson et al. (2012) find the λ that minimizes the variance of the difference between y_j and the linear rule $\hat{y}_j = \lambda z_j + (1-\lambda)\hat{\psi}x_j$. Yet $var(y_j - \hat{y_j}) = E[(y_j - \hat{y_j})^2]$ by the unbiasedness of z_j and the properties of OLS fitted values. In other words, Henderson et al. (2012) restrict attention to linear combinations of z_j and $\hat{\psi}x_j$ and then derive the optimal $\hat{y}_j$ using a squared-error loss function. Efron and Morris (1973) show that the optimal λ from this procedure yields the "linear Bayes rule" which gives the Bayesian optimal weighting regardless of the actual underlying distributions.

in a few intuitively appealing steps. Second, it suggests treating cities with and without existing growth estimates in a unified manner: cities without existing estimates can be viewed as draws from the prior with σ_z^2 tending to infinity. Third, it illustrates the robustness of the Henderson et al. (2012) procedure to misspecification.

2.2 Robustness of the Henderson et al. (2012) Weighting

The previous discussion suggests the robustness of the Henderson et al. (2012) weighting. In order to make this statement more precise, note that $\min_{\lambda}\{E[(\lambda(z_j - y_j) + (1-\lambda)(\psi x_j - y_j))^2]\}$ is equivalent to $\min_{\lambda}\{\lambda^2\sigma_z^2 + (1-\lambda)^2 E[(y_j - \psi x_j)^2]\}$ because $cov(\epsilon_{xj}, \epsilon_{zj}) = 0$ and city growth is measured with classical measurement error. This minimization problem yields $\lambda^* = \frac{E[(y_j-\psi x_j)^2]}{E[(y_j-\psi x_j)^2]+\sigma_z^2}$ and corresponding MSE $\frac{E[(y_j-\psi x_j)^2]\sigma_z^2}{E[(y_j-\psi x_j)^2]+\sigma_z^2} \leq \frac{\sigma_y^2\sigma_z^2}{\sigma_y^2+\sigma_z^2} < \sigma_z^2$.

This simple proof demonstrates that the MSE corresponding to λ^* decreases as the R^2 of the regression of y on x increases. The MSE is minimized, in turn, when $\psi x_j = E[y_j|x_j]$.[12]

Consequently, the optimal λ in an environment where the conditional mean is misspecified will generate a composite estimate with lower MSE than that of z_j taken on its own, although this estimate will have higher MSE than that calculated using $E[y_j|x_j]$. In other words, specification error reduces the weight on the OLS fitted values in the composite estimate because the fitted values do a poorer job predicting city growth. This illustrates how the validity of the Henderson et al. (2012) weighting rests on the assumption that $cov(\epsilon_{xj}, \epsilon_{zj}) = 0$.[13] It does not rest on the exogeneity of x_j or on the ability of the researcher to specify the exact relationship between x_j and y_j (although correct specification will lead to better estimates). It does not even depend on the ability of x_j to predict y_j: for example as author measurement error grows to infinity, the shrinkage estimator continues to be valid, eventually reducing to the average of observed city growth and its overall mean.[14]

In sum, the optimal weighting in equation (4) produces improved estimates even if normality is violated, x_j is endogenous, the regression is misspecified and x_j is measured with error. This discussion also illustrates how $\hat{y}_j$ is guaranteed to be an improvement over z_j only in a global sense: individual $\hat{y}_j$ may be worse estimates of y_j than z_j (e.g. Efron and Morris, 1971, 1972).

[12] This can be seen by noting that $E[(y_j - \psi x_j)^2] = E[(y_j - E[y_j|x_j])^2] + E[(\psi x_j - E[y_j|x_j])^2]$.

[13] Because the assumption of classical measurement error is commonplace I abstract from the consequences of violations of this assumption.

[14] For an intuitive illustration of this estimator see Efron and Morris (1977).

2.3 Summing Across Cities and Years

Here I briefly discuss how summing across years and between cities can reduce measurement error.

Begin with summing author counts within cities over a window of length w around century t. Assuming that π_t is constant over this window, $\sum_{s\in[-w/2,w/2)} X^*_{jt+s}$ is approximately distributed $N(\pi_t \sum_{s\in[-w/2,w/2)} C_{jt+s}, \pi_t(1-\pi_t)\sum_{s\in[-w/2,w/2)} C_{jt+s})$. The delta method, in turn, implies that $ln(\sum_{s\in[-w/2,w/2)} X^*_{jt+s})$ is approximately distributed $N(ln(\pi_t)+ln(\sum_{s\in[-w/2,w/2)} C_{jt+s}), \frac{1-\pi_t}{\pi_t \sum_{s\in[-w/2,w/2)} C_{jt+s}})$. Consequently, the distribution of $ln(\frac{\sum_{s\in[-w/2,w/2)} X^*_{jt+s}}{\sum_{s\in[-w/2,w/2)} X^*_{jt-100+s}})$ is centered around $ln(p_t) - ln(p_{t-100}) + ln(\frac{\frac{\sum_{s\in[-\frac{w}{2},\frac{w}{2})} C_{jt+s}}{w}}{\frac{\sum_{s\in[-\frac{w}{2},\frac{w}{2})} C_{jt-100+s}}{w}})$.

In other words, the difference between the sum of authors within a window w around t and t-100 can be interpreted as an estimate of the change in the probability an author's works survive to the present plus the growth rate of the average number of scholars working in a city within the window. Furthermore, it becomes increasingly probable that $\sum_{s\in[-\frac{w}{2},\frac{w}{2})} C_{jt+s}$ is large as w grows, helping to alleviate concerns regarding the use of the normal approximation to the binomial distribution.[15]

Similar arguments apply to summing author counts across a group of cities denoted by the indicator d_j. As the number of these cities becomes large the approximation $ln(\sum_{d_j=1} X^*_{jt}) = ln(\pi_t) + ln(\sum_{d_j=1} C_{jt})$ improves.

3 Empirical Analysis

The conceptual framework developed in the previous section suggests estimating a regression of the form:

$$z_{jt} = \alpha_t + \psi_t x_{jt} + \xi_{jt} \tag{5}$$

where z_{jt} denotes existing measurements of city growth and x_{jt} denotes author growth in city j and century t.[16] Year fixed effects control for changes in the probability of observing authors across

[15]This discussion also illustrates the potential problems involved when including cities with low values of $\pi_t \sum_{s\in[-w/2,w/2)} C_{jt+s}$ in the analysis. For expositional clarity I abstract from these issues and use the logarithm as one plus the sum of authors as my measure of author counts. From an empirical standpoint this abstraction is justified by the fact that the change in the logarithm of one plus author growth is robustly correlated with city growth even for cities with small numbers of observed author deaths.

[16]For regressions at the century level, x_{jt} is the difference between the logarithm of one plus the sum of all authors who died on the 100 year interval [t-50,t+50) and the logarithm of one plus the sum of those who died on the interval

centuries as well as other time-varying factors.[17]

Figure 4 presents scatter plots of the relationship between z_{jt} and x_{jt} by century. Dashed lines provide the OLS fitted values in these graphs whereas the solid line gives the fitted values from a LOWESS smoother with bandwidth 0.8. The general linear nature of the relationship between the two variables is evident, providing visual evidence that linear regression provides a reasonable approximation to the conditional expectation function.

Table two confirms this point more formally.[18] In this table, I estimate equation 5 using the 1,843 Bairoch et al. (1988) growth rates. For the sake of clarity, I begin by estimating this equation separately by century. Columns 1-5 show that aside from the 1700 coefficient, the estimated δ_t are roughly stable.[19] Without further assumptions, it is impossible to determine whether the relatively large 1700 coefficient is due to a drop in measurement error or structural changes.[20]

The row labeled Linearity provides the p-value from Ramsey (1969)'s RESET test of this specification. The data are generally consistent with linearity, suggesting that the composite estimates resulting from the regression fitted values will be close to optimal.

Columns 6 and 7 present estimates pooling the data. Column 6 presents the variance-weighted average of the century coefficients (i.e. I restrict ψ_t to be constant across centuries in equation 5). In column 7 I add a city specific time trend to equation 5 by including city fixed effects. This specification shows that variation in author growth predicts variation in city growth about their respective growth paths, casting doubt on the possibility that author and city growth are spuriously related.

As table two makes clear, each century uses a different set of cities to estimate the coefficients of interest due to missing data in Bairoch et al. (1988). The conceptual framework, however, suggests that this selection is not likely to severely affect the analysis. To see this, note that city growth is independent

[t-150,t-50). In the online appendix I provide empirical justification for this choice as well as showing that the qualitative implications of the results are largely robust to varying this window. Throughout I use the log difference of city populations to estimate the city growth rate.

[17]In the appendix, I show that the results are robust to the use of country-time fixed effects, helping to alleviate concerns that library-specific collection patterns drive the results.

[18]I report heteroscedasticity-robust standard errors unless otherwise noted. Although city growth is believed to be i.i.d. (e.g. Eeckhout, 2004), the conceptual framework suggests that measurement error in author growth will introduce heteroscedasticity. The results, however, are robust to the use of both homoscedastic and HAC standard errors.

[19]The p-value associated with the null hypothesis that these coefficients are equal is 0.57.

[20]The most obvious such structural change would be a change in relationship between author counts and city sizes between 1600 and 1700. Such a change, however, is impossible to empirically accommodate within the theoretical framework since the β_{jt} are not identified. Nevertheless, because author growth continues to predict city growth the empirical Bayes setup remains relevant as discussed above.

of the initial level by Gibrat's law (e.g. Eeckhout, 2004) which implies that $E[z_{jt}|Y_{jt} > c] = E[y_{jt}]$. Consequently, inasmuch as selection is related to initial city size, the composite estimates will have lower mean-squared error regardless of how selection affects the regression coefficients.

To derive the composite estimates, it is necessary to obtain estimates of the signal-to-noise ratio $\phi_t = \frac{\sigma^2_{yt}}{\sigma^2_{yt}+\sigma^2_{zt}}$. I use the Bairoch et al. (1988) growth estimates and those provided by de Vries (1989) to obtain the estimates $\hat{\phi}_{1700} = .885$ and $\hat{\phi}_{1600} = .784$.[21] Subbing these values and $\hat{\rho}_{xzt}$ into equation 4 yields (assuming the signal-to-noise ratio is constant prior to 1600) $\hat{\lambda}^*_{1700} = .86683792$, $\hat{\lambda}^*_{1600} = .77580613$, $\hat{\lambda}^*_{1500} = .77293622$, $\hat{\lambda}^*_{1400} = .77275342$ and $\hat{\lambda}^*_{1300} = .77059573$. For years prior to 1300 I set $\hat{\lambda}^* = 0$ and the regression coefficients to their values in 1300.

In table 3 I illustrate how the empirical Bayes weighting leads to new city growth estimates for Nantes, France. In the first column I provide the century whereas the second column provides the Bairoch et al. (1988) city size estimates (in thousands). The third column presents the $z_{Nantes,t}$, column 4 one plus the number of author deaths and column 5 the $x_{Nantes,t}$. In column 6, I provide estimates of the signal-to-noise ratio and column 7 provides the estimated correlation between city and author growth. Column 8 provides estimates of the optimal weight on existing estimates and column 9 provides the fitted values estimated from equation 5. Column 10 provides the composite estimate and column 11 provides one population time path that is consistent with the composite estimates.[22]

The data corresponding to 1600 illustrate the shrinkage estimator in action. The existing growth estimate of 0.58 in column 3 contrasts with the OLS estimate of 0.24 in column 9. The optimal λ incorporates this information by pulling the existing estimate towards the OLS estimate. To do this, the shrinkage estimator weights the existing estimate by 0.78 leading to the composite growth estimate of 0.50.

This table also demonstrates that growth estimates are likely less reliable prior to 1300 since relatively small author counts combine with the impossibility of estimating regression coefficients during these centuries. Nevertheless, given the complete lack of estimates for most cities during this period I

[21] De Long and Shleifer (1993, p. 676) imply that these estimates are largely independently obtained. Yet even if they are not (and thus rely on at least some of the same sources) my estimates of the signal to noise ratio will be too large, placing higher weight on existing growth estimates. In other words, failure of this assumption biases the new estimates towards existing growth estimates.

[22] This time path is derived by assuming that the Bairoch et al. (1988) estimate for 1700 is correct and then iterating the population level backwards using the composite estimates.

provide these estimates.[23]

I perform a similar exercise for the remaining 1,320 Bairoch et al. (1988) cities outside southeastern Europe.[24] This leads to 13,210 new or modified population estimates on the interval [800,1700] as compared to the 3,020 estimates provided by Bairoch et al. (1988). Combining these new data with the existing Bairoch et al. (1988) data yields a new dataset with 18,910 population estimates on the interval [800,1850].

3.1 Authors and City Growth in the MENA Region

Here I briefly explore the generality of the relationship between author and city growth by extending the analysis from Europe to authors in the MENA region.[25] Given the relatively small numbers of MENA authors in the VIAF data I view this analysis as preliminary and exploratory.[26] My aim is to provide suggestive evidence that the relationship between city and author growth extends beyond Europe in order to motivate future research.

I begin from the Bosker et al. (2013) growth estimates for cities in the MENA region. I then pool all observations to estimate regression 5 on the pre-1700 data. The broken line in Figure 5 plots the resulting relationship (net of century dummies) whereas the solid line provides the fitted values from an identical regression pooling the pre-1700 European data.[27] The European and MENA slopes are strikingly similar, suggesting that the relationship between city size and author growth is not limited to Europe.

[23]Note also that author counts in 800 are calculated using only the interval [800,850). For expositional purposes I abstract from this asymmetry which has minimal empirical implications.

[24]To derive estimates of population levels I proceed iteratively. If Bairoch et al. (1988) provide an estimate for 1700 I use that estimate as in the Nantes case. If there is not an estimate for 1700, I use the 1600 estimate to derive the population time path. If there is no estimate for 1700 and 1600 I use the 1500 estimate and so forth.

[25]To be precise, the previous regression analysis already contained a handful of MENA cities primarily from the Iberian Peninsula.

[26]Of the total 243,451 geo-referenced authors only 13,693 died within MENA boundaries. This relative lack of data does not seem to stem from VIAF bias but from the fact that a large number of historical works produced in the MENA region are unknown (e.g. Saliba, 2002).

[27]I limit analysis to the pre-1700 data since there is evidence of a jump in the relationship between author and city growth in 1700 as discussed above.

4 The Rise of the Atlantic Trade

How do the composite estimates improve upon existing data? To explore this question I revisit Acemoglu et al. (2005)'s influential study. This article used the Bairoch et al. (1988) data to provide evidence that cities engaged in Atlantic trade grew more rapidly than those that did not after 1500. In many ways, this empirical environment is ideal to demonstrate the value of the new dataset since the original study was limited by missing data. For example, Acemoglu et al. (2005) were forced to drop cities involved in Atlantic trade such as Nantes, Rotterdam, Liverpool and Cadiz from their balanced sample because city size estimates were not available for some centuries (Acemoglu et al., 2002, p. 58). The new dataset provides estimates of city growth for these cities as well as others that Acemoglu et al. (2005) were forced to omit from their analysis.

I begin by replicating column 2 from table 5 of Acemoglu et al. (2005, p. 560) in column 1 of table 4. Throughout this table I limit the sample to the years 1300 and after, reporting homoscedastic standard errors for comparison with the original Acemoglu et al. (2005) study. In column 2, I present results using one plus the logarithm of author counts as the dependent variable on the Acemoglu et al. (2005) sample of cities. In column 3, I use the new estimates of city sizes on the Acemoglu et al. (2005) sample while column 4 expands the dataset to the balanced sample of cities in the new dataset (in these data the 4 additional Atlantic trading cities are added to the original group of 13).[28]

The results in columns 3 and 4 are dependent on city population levels which are not pinned down by the statistical procedure described above.[29] To obtain results that are not dependent on levels, in columns 6-9 I report estimates of the γ_t from a regression of the form (here that estimated in column 7):

$$z_{jt} = \alpha_t + \beta_t \cdot WE_j + \sum_{t \in \{1400,1500,1600\}} \gamma_t \cdot atl_j \cdot (d_t - d_{t+100}) + \gamma_{1700} \cdot atl_j \cdot d_{1700} + \epsilon_{jt} \tag{6}$$

where z_{jt} again denotes the relevant log difference in city j between century t and t-100, WE_j is an indicator equal to one if city j is in western Europe and atl_j is an indicator equal to one if Acemoglu et al. (2005) define a city to be an Atlantic trader.

[28] This new balanced panel is composed of 963 cities compared with the 193 cities in the original balanced panel.

[29] In other words, there are infinite time paths that are consistent with the revised growth estimates although for expositional simplicity I have assumed that the Bairoch et al. (1988) estimates are correct for 1700 to derive one level path that is consistent with the growth estimates as explained above.

Column 6 reports coefficients of interest for the Acemoglu et al. (2005) sample and data whereas column 5 estimates the corresponding regression in levels for comparison. In column 7 I report results obtained using the log difference of author counts as the dependent variable. In column 8 I report results obtained using the composite data on the Acemoglu et al. (2005) sample whereas column 9 presents results on the balanced sample of cities in the new dataset. Throughout table 4, the standard errors using the composite data are noticeably smaller than those in the original Acemoglu et al. (2005) regression, illustrating the efficiency gains of the composite estimates.

When exactly did the Atlantic traders begin to diverge from the rest of Western Europe? Although the absence of yearly city size estimates made this question unanswerable in the original study, the high-frequency nature of author counts allows for a preliminary analysis of this question. To do this, I constructed differences-in-differences coefficients by year by summing all author deaths in a given year within Atlantic and non-Atlantic cities and using the fact that $ln(\sum_{d_j=1} X_{jt})$ is approximately equal to $ln(\pi_t) + ln(\sum_{d_j=1} C_{jt})$. This relationship makes estimating $[ln(E(C_{jt}|d_j = 1, t = t) - ln(E(C_{jt}|d_j = 1, t = t^*)] - [ln(E(C_{jt}|d_j = 0, t = t) - ln(E(C_{jt}|d_j = 0, t = t^*)]$ straightforward. Figure 6 graphs these year-by-year estimates (with t^* set to 1425) and then smooths these estimates.[30] This figure provides suggestive evidence that the divergence between the Atlantic traders and the rest of Western Europe began in the decades following Columbus' Atlantic crossing.

5 Conclusion

This paper has provided evidence that library holdings can be used to measure historical economic activity. In particular, changes in the number of authors affiliated with a given city are strongly and robustly correlated with existing city growth estimates. I derived the optimal weighting in Henderson et al. (2012) within an empirical Bayes framework and then used this weighting to provide new and improved city growth estimates. The paper illustrated the value of the new dataset by revisiting the rise of the Atlantic traders, confirming the results in Acemoglu et al. (2005) and shedding greater light on the exact timing of the divergence between these cities and those in the remainder of western Europe.

[30] To derive this graph I began from the 963 cities in the balanced panel, dropped cities in the countries Acemoglu et al. (2005) define as belonging to eastern Europe and summed the remaining author counts by year in the Atlantic trader and non-Atlantic trader cities. The differences-in-differences coefficients are smoothed using a LOWESS smoother with bandwidth 0.1.

As with luminosity data, author counts will be of great use in non-Western regions where historical estimates of city sizes are imprecise or unavailable. This observation suggests the value of both expanding the MENA dataset as well as extending the analysis to other non-European regions such as East Asia. It is my hope that future research will both refine and exploit this novel proxy to enhance our understanding of economic development.

References

Acemoglu, D., S. Johnson, and J. Robinson, "The Rise of Europe: Atlantic Trade, Institutional Change and Economic Growth," *NBER Working Paper 9378*, 2002.

_ , _ , and _ , "The Rise of Europe: Atlantic Trade, Institutional Change, and Economic Growth," *American Economic Review*, 2005, *95* (3), 546–579.

Ades, A. and E. Glaeser, "Trade and Circuses: Explaining Urban Giants," *The Quarterly Journal of Economics*, 1995, *110* (1), 195–227.

Angrist, J., P. Hull, and P. Pathak, "Leveraging Lotteries for School Value-Added: Testing and Estimation," *Quarterly Journal of Economics*, 2017, *132* (2), 871–919.

Bairoch, P., J. Batou, and P. Chevre, *La population des ville europeenees de 800 a 1850: Banque de donnees et analyse sommaire de resultats.*, Geneva: Librarie Droz, 1988.

Barjamovic, G., T. Chaney, K. Cosar, and A. Hortacsu, "Trade, Merchants, and the Lost Cities of the Bronze Age," *Quarterly Journal of Economics*, 2019, pp. 1455–1503.

Baten, J. and J.L. van Zanden, "Book production and the onset of modern economic growth," *Journal of Economic Growth*, 2008, *13*, 217–235.

Bosker, M., E. Buringh, and J.L. van Zanden, "From Baghdad to London: Unraveling Urban Development in Europe, the Middle East and North Africa, 800-1800," *Review of Economics and Statistics*, 2013, *95* (4), 1418–1437.

Buringh, E. and J.L. van Zanden, "Charting the "Rise of the West": Manuscripts and Printed Books in Europe, A Long-Term Perspective from the Sixth through Eighteenth Centuries," *Journal of Economic History*, 2009, *69* (2), 409–445.

Chetty, R., J.N. Friedman, and J.E. Rockoff, "Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates," *American Economic Review*, 2014, *104* (9), 2593–2632.

Cummins, N., "Lifespans of the European Elite, 800-1800," *Journal of Economic History*, 2017, *77* (2), 406–439.

de la Croix, D. and O. Licandro, "The longevity of famous people from Hammurabi to Einstein," *Journal of Economic Growth*, 2015, *20*, 263–303.

_ , **F. Docquier, A. Fabre, and R. Stelter**, "The Academic Market and the Rise of Universities in Medieval and Early Modern Europe (1000-1800)," *Working Paper*, 2019.

De Long, J.B. and A. Shleifer, "Princes and Merchants: European City Growth before the Industrial Revolution," *The Journal of Law & Economics*, 1993, *36* (2), 671–702.

de Vries, J., "Review of La population des villes europeennes," *Journal of Economic History*, 1989, *49* (4), 1007–1008.

Dittmar, J., "Information Technology and Economic Change: the Impact of the Printing PRess," *Quarterly Journal of Economics*, 2011, *126* (3), 1133–1172.

_ **and S. Seabold**, "New Media and Competition: Printing and Europe's Transformation after Gutenberg," *Working Paper*, 2019.

Eeckhout, J., "Gibrat's Law of (All) Cities," *American Economic Review*, 2004, *94* (5).

Efron, B. and C. Morris, "Limiting the Risk of Bayes and Empirical Bayes Estimators–Part I: The Bayes Case," *Journal of the American Statistical Association*, 1971, *66* (336), 807–815.

_ **and** _ , "Limiting the Risk of Bayes and Empirical Bayes Estimators–Part II: The Empirical Bayes Case," *Journal of the American Statistical Association*, 1972, *67* (337), 130–139.

_ **and** _ , "Stein's Estimation Rule and Its Competitors–An Empirical Bayes Approach," *Journal of the American Statistical Association*, 1973, *68* (341), 117–130.

_ **and** _ , "Stein's Paradox in Statistics," *Scientific American*, 1977, *236* (5), 119–127.

Fay III, R.E and R.A. Herriot, "Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data," *Journal of the American Statistical Assocation*, 1979, *74* (366), 269–277.

Fessler, P. and M. Kasy, "How to use Economic Theory to Improve Estimators: Shrinking toward Theoretical Restrictions," *Review of Economics and Statistics*, 2019, *101* (4), 681–698.

Gergaud, O., M. Laouenan, and E. Wasmer, "A Brief History of Human Time," *Working Paper*, 2017.

Glaeser, E., A. Saiz, G. Burtless, and W. Strange, "The Rise of the Skilled City," *Brookings-Wharton Papers on Urban Affairs*, 2004, pp. 47–105.

Gomez-Lievano, A., H. Youn, and L.M.A Bettencourt, "The Statistics of Urban Scaling and Their Connection to Zipf's Law," *PLoS ONE*, 2012, *7* (7).

Henderson, J.V., A. Storeygard, and D. Weil, "Measuring Economic Growth from Outer Space," *American Economic Review*, 2012, *102* (2), 994–1028.

Hickey, T. and J. Toves, "Managing Ambiguity in VIAF," *D-Lib Magazine*, 2014, *20* (7/8).

Kane, T. and D. Staiger, "Estimating Teacher Impacts on Student Achievement: An Experimental Evalution," *NBER Working Paper*, 2008, (14607).

Katz, D., J. Baptista, S.P. Azen, and M.C. Pike, "Obtaining Confidence Intervals for the Risk Ratio in Cohort Studies," *Biometrics*, 1978, *34* (3), 469–474.

Koopman, P.A.R., "Confidence Intervals for the Ratio of Two Binomial Proportions," *Biometrics*, 1984, *40* (2), 513–517.

McLeod, T., "Magdeburg, sack of (1631)," in R. Holmes, C. Singleton, and S. Jones, eds., *The Oxford Companion to Military History*, Oxford: Oxford University Press, 2001.

Michalopoulos, S. and E. Papaioannou, "Spatial Patterns of Development: a Meso Approach," *Annual Reviews of Economics*, 2018, *10*, 383–410.

Mokyr, J., "Urbanization, Technological Progress, and Economic History," in H. Giersch, ed., *Urban agglomeration and economic growth*, Berlin: Springer-Verlag, 1993, pp. 3–38.

Morris, C., "Parametric Empirical Bayes Inference: Theory and Applications," *Journal of the American Statistical Association*, 1981, *78* (381), 47–55.

Mullainathan, S. and J. Spiess, "Machine Learning: an Applied Econometric Approach," *Journal of Economic Perspectives*, 2017, *31* (2), 87–106.

Nunn, N. and N. Qian, "The Potato's Contribution to Population and Urbanization: Evidence from a Historical Experiment," *Quarterly Journal of Economics*, 2011, *126* (2), 593–650.

Ramsey, J., "Tests for Specification Errors in Classical Linear Least Squares Regression Analysis," *Journal of the Royal Statistical Society*, 1969, *31* (2), 350–371.

Rollo-Koster, J., *Avignon and Its Papacy, 1309-1417*, London: Rowman & Littlefield, 2015.

Saliba, G., "Greek Astronomy and the Medieval Arabic Tradition," *American Scientist*, 2002, *90* (4), 360–367.

Schich, M., C. Song, Y. Ahn, A. Mirsky, M. Martino, and A. Barabasi, "A network framework of cultural history," *Science*, 2014, *345*, 558–562.

Zutshi, P.N.R., "The Avignon Papacy," in M. Jones, ed., *The New Cambridge Medieval History VI*, Cambridge: Cambridge Univerity Press, 2000, pp. 653–673.

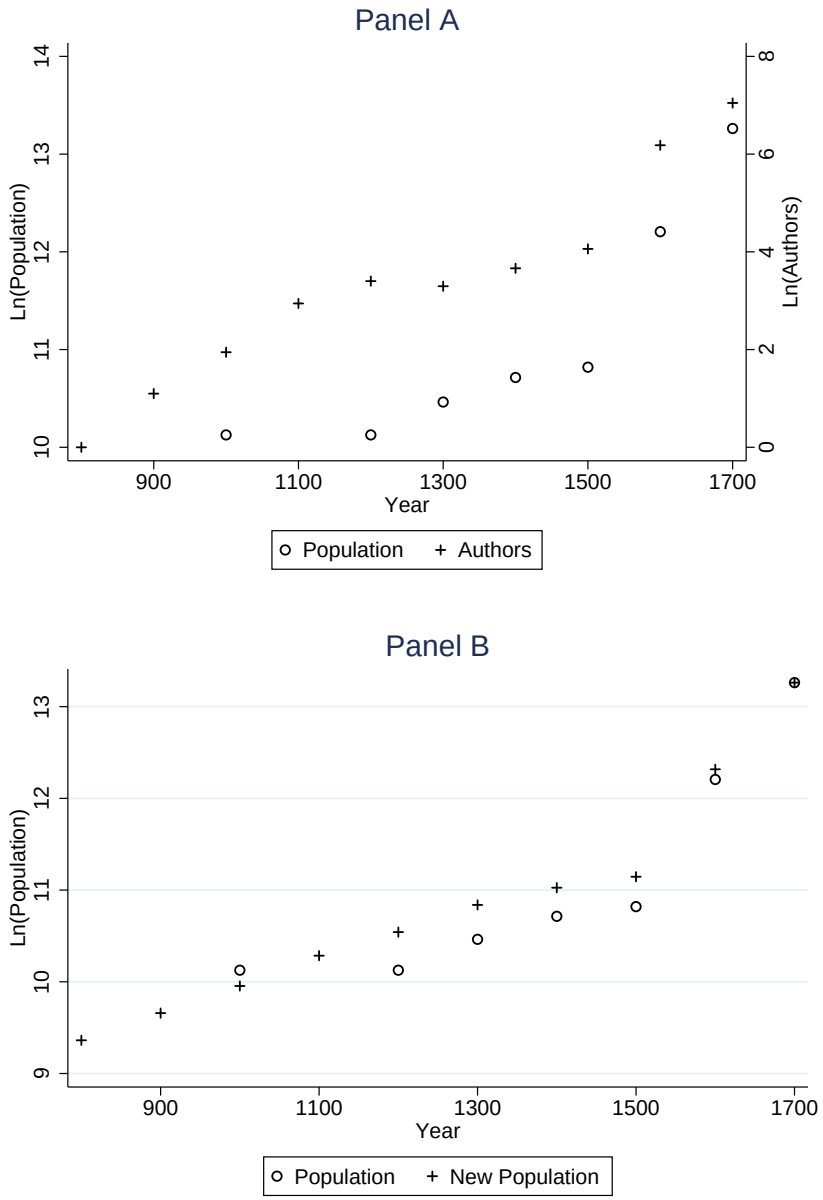

Figure 1: **Population of London and London-based Authors by Century**

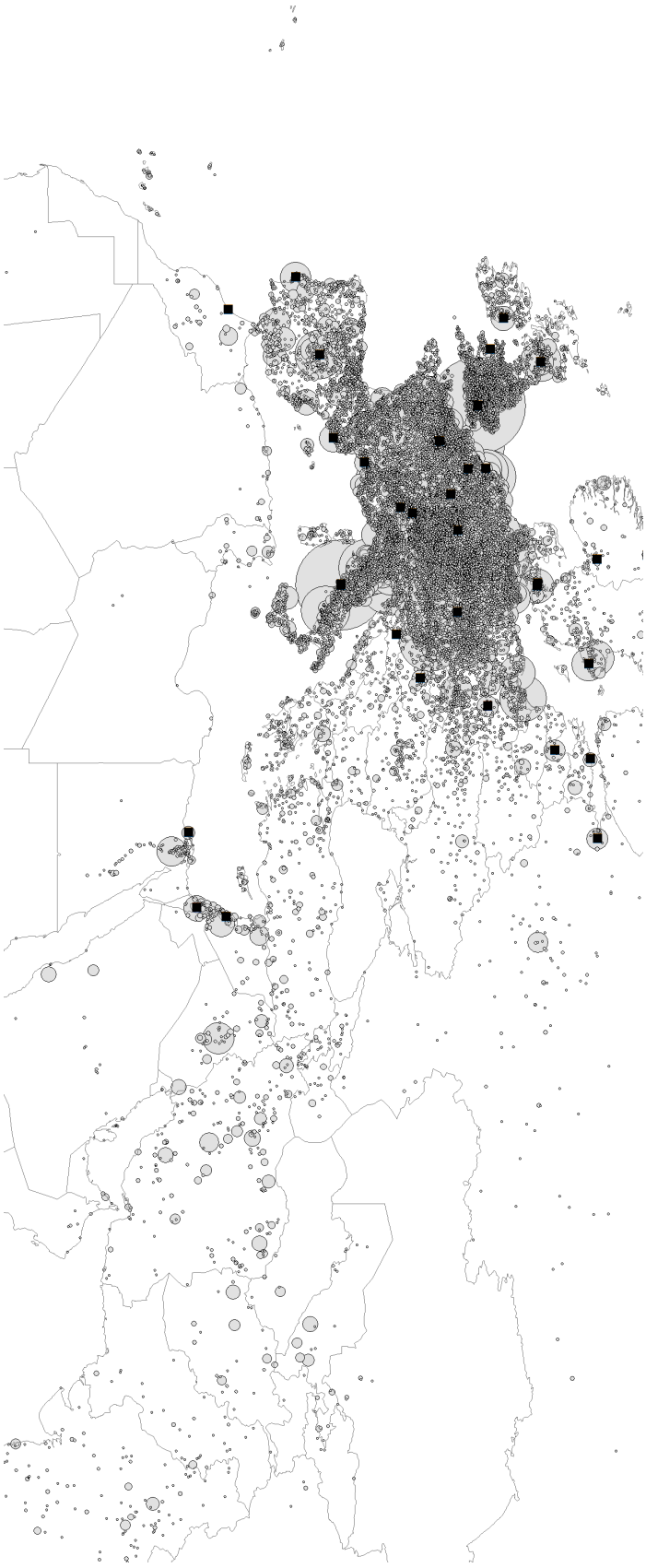

Figure 2: **Author Death Places and VIAF Libraries**
Grey circles provide the number of author deaths in a city and black squares represent the location of the VIAF libraries.

Table 1: **Author Counts and Population Estimates for Selected Cities**

	London	Paris	Venice	Vienna	Madrid	Rome	Nantes	Baghdad	Cordoba
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Panel A: Authors								
800	0	7	5	1	0	11	0	106	6
900	2	15	7	2	3	41	0	135	27
1000	6	7	12	1	5	42	1	94	48
1100	18	19	7	3	4	49	2	72	22
1200	29	114	21	21	9	83	2	54	25
1300	51	175	31	24	14	97	0	27	5
1400	56	158	168	55	12	167	5	13	8
1500	160	464	417	140	55	576	9	10	33
1600	968	1528	782	298	399	1260	17	16	73
1700	2203	3614	503	795	489	1435	35	12	38
	Panel B: Existing Population Estimates								
800	-	25	-	-	-	50	-	350	169
900	-	-	7	-	-	40	-	450	-
1000	25	20	45	-	-	35	-	300	450
1100	-	-	-	-	-	-	-	250	-
1200	25	110	70	12	-	35	-	200	60
1300	35	150	110	20	3	30	8	95	60
1400	45	275	100	20	-	33	-	90	40
1500	50	225	100	20	13	55	14	60	35
1600	200	300	151	50	65	100	25	35	31
1700	575	500	138	114	140	135	40	30	31

Notes: panel A provides the number of authors in the VIAF data that died in a given city on the interval [t-50,t+50). Panel B lists the available Bairoch et al. (1988); Bosker et al. (2013) population estimates (in thousands of inhabitants) for each city by century.

Table 2: **City and Author Growth by Century**

	Δln(Population)						
	1700	1600	1500	1400	1300	All	All
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\Delta ln(Authors)$	0.32***	0.12***	0.16***	0.21***	0.19**	0.21***	0.16***
	(0.03)	(0.03)	(0.04)	(0.06)	(0.09)	(0.02)	(0.03)
Linearity	[0.10]	[0.12]	[0.20]	[0.25]	[0.37]	-	-
N	728	466	271	261	117	1843	1843
R^2	0.14	0.04	0.05	0.05	0.06	0.11	0.51
Time Dummies	-	-	-	-	-	Yes	Yes
City Dummies	-	-	-	-	-	No	Yes

Notes: the first five columns provide the $\hat{\psi}$ from the regression $z_j = \alpha + \psi x_j + \xi_j$ estimated by century. In this regression z_j denotes existing population growth estimates and x_j author growth. The row labelled Linearity provides the p-value from Ramsey (1969)'s RESET test for linearity. Columns 6 and 7 provide the $\hat{\psi}$ obtained from pooling the data and estimating $z_{jt} = \beta_j + \alpha_t + \psi x_{jt} + \xi_{jt}$. Heteroscedasticity-robust standard errors in parentheses. ***, **and *denote significance at the 1%, 5% and 10% levels.

Table 3: **Comparison of Existing and New Growth Rates: Nantes, France**

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Year	City Level	z	Author Level+1	x	$\hat{\phi}$	$\hat{\rho}_{xz}$	$\hat{\lambda}^*$	$\hat{\alpha}+\hat{\psi}x$	$\hat{y}$	New Level
800	-	-	1	-	-	-	-	-	-	4.42
900	-	-	1	0	-	-	0	0.19	0.19	5.35
1000	-	-	2	0.41	-	-	0	0.25	0.25	6.88
1100	-	-	3	0.29	-	-	0	0.23	0.23	8.70
1200	-	-	3	0	-	-	0	0.19	0.19	10.52
1300	8	-	1	-0.69	0.78	0.24	0	0.08	0.08	11.45
1400	-	-	6	1.25	0.78	0.22	0	0.16	0.16	13.49
1500	14	-	10	0.45	0.78	0.22	0	0.16	0.16	15.75
1600	25	0.58	18	0.55	0.78	0.19	0.78	0.24	0.50	26.05
1700	40	0.47	36	0.67	0.89	0.37	0.87	0.16	0.43	40

Notes: column 2 provides the Bairoch et al. (1988) population estimates where available and column 3 provides the corresponding growth rates. Columns 4 and 5 detail one plus the number of authors who died on the interval [t-50,t+50) and the growth rate of this quantity. Column 6 reports the signal to noise ratio estimated from the Bairoch et al. (1988) and de Vries (1989) data whereas column 7 provides the sample correlation between city and author growth. Column 8 reports the optimal weight on the existing growth estimates and column 9 provides the OLS growth estimates. Column 10 gives the composite estimate corresponding to the optimal weighting and column 11 provides the city level time path assuming the Bairoch et al. (1988) 1700 estimate is correct.

Table 4: **Atlantic Trade**

	Old	Aut	New	New	Old	Old	Aut	New	New
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Western dummies p-value	[0.05]	[0.00]	[0.02]	[0.05]	[0.04]	[0.01]	[0.07]	[0.00]	[0.03]
Atlantic Port x 1400					0.18	0.18	-0.08	0.14	0.19*
					(0.23)	(0.14)	(0.21)	(0.12)	(0.10)
Atlantic Port x 1500	-0.05	0.45*	-0.03	0.02	0.05	0.05	0.35	0.05	0.12
	(0.20)	(0.27)	(0.18)	(0.14)	(0.23)	(0.19)	(0.29)	(0.17)	(0.15)
Atlantic Port x 1600	0.46**	1.05***	0.39**	0.45***	0.55**	0.55**	0.99***	0.46**	0.52***
	(0.20)	(0.27)	(0.18)	(0.14)	(0.23)	(0.23)	(0.36)	(0.20)	(0.18)
Atlantic Port x 1700	0.62***	1.21***	0.53***	0.88***	0.71***	0.71***	1.15***	0.60**	0.75***
	(0.20)	(0.27)	(0.18)	(0.14)	(0.23)	(0.27)	(0.42)	(0.24)	(0.21)
Atlantic Port x 1750	0.71***		0.62***	0.96***	0.80***	0.80***		0.69***	0.77***
	(0.20)		(0.18)	(0.14)	(0.23)	(0.30)		(0.26)	(0.23)
Atlantic Port x 1800	0.92***		0.83***	1.13***	1.01***	1.01***		0.90***	0.91***
	(0.20)		(0.18)	(0.14)	(0.23)	(0.33)		(0.29)	(0.25)
Atlantic Port x 1850	1.00***		0.91***	1.26***	1.09***	1.09***		0.98***	1.03***
	(0.20)		(0.18)	(0.14)	(0.23)	(0.36)		(0.31)	(0.27)
R^2	0.79	0.86	0.80	0.81	0.79	0.08	0.14	0.10	0.13
N	1544	965	1544	7704	1544	1351	772	1351	6741
Sample	AJR	AJR	AJR	All	AJR	AJR	AJR	AJR	All
Spec	Level	Level	Level	Level	Level	Dif	Dif	Dif	Dif

Notes: Column headings denote the source of the dependent variable. Old denotes the original Bairoch et al. (1988) population estimates, Aut the number of authors and New the composite estimates. The row Sample details the sample on which the regressions are run: AJR denotes the Acemoglu et al. (2005) balanced panel whereas All denotes the balanced panel in the composite data. The row Spec provides whether the regressions are run in city levels or first differences. Column 1 replicates column 2 from table 5 of Acemoglu et al. (2005). Column 2 presents output from the Acemoglu et al. (2005) specification using the logarithm of authors as the dependent variable. In column 3 I provide coefficients from the Acemoglu et al. (2005) specification with the composite (level) data as the dependent variable. In column 4 I expand the dataset to include the balanced panel of cities in the composite (level) data. Column 5 again replicates column 2 from table 5 of Acemoglu et al. (2005) adding a 1400 interaction to facilitate working in first-differences. Columns 6-9 work in first differences but otherwise follow the pattern of columns 1-4. Homoscedastic standard errors in parentheses. ***, **and *indicate significance at the 1%, 5% and 10% levels.

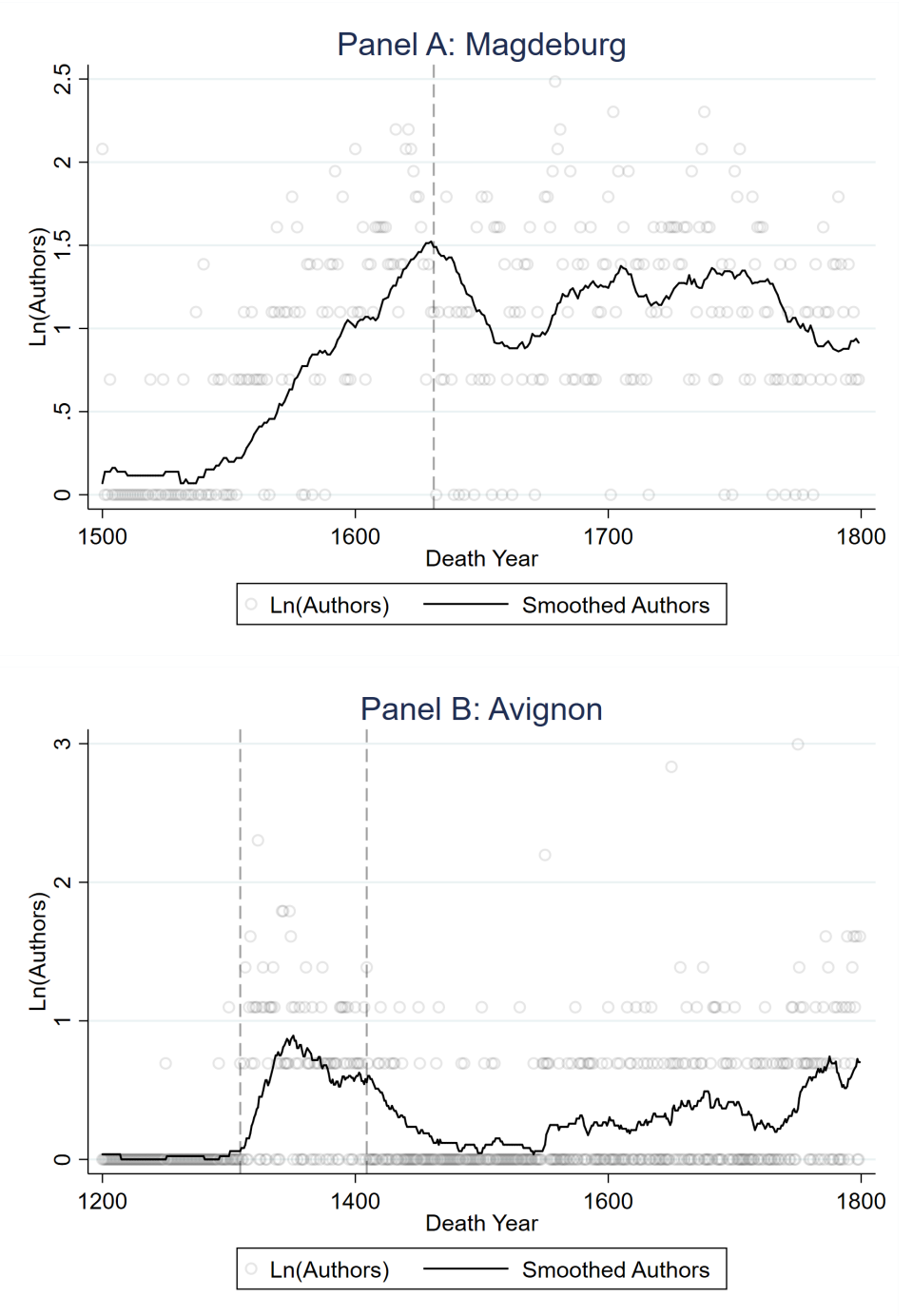

Figure 3: **Events and Author Counts in Magdeburg, Germany and Avignon, France** Vertical line in Panel A denotes Magdeburg Massacre whereas the lines in Panel B delimit the Avignon Papacy

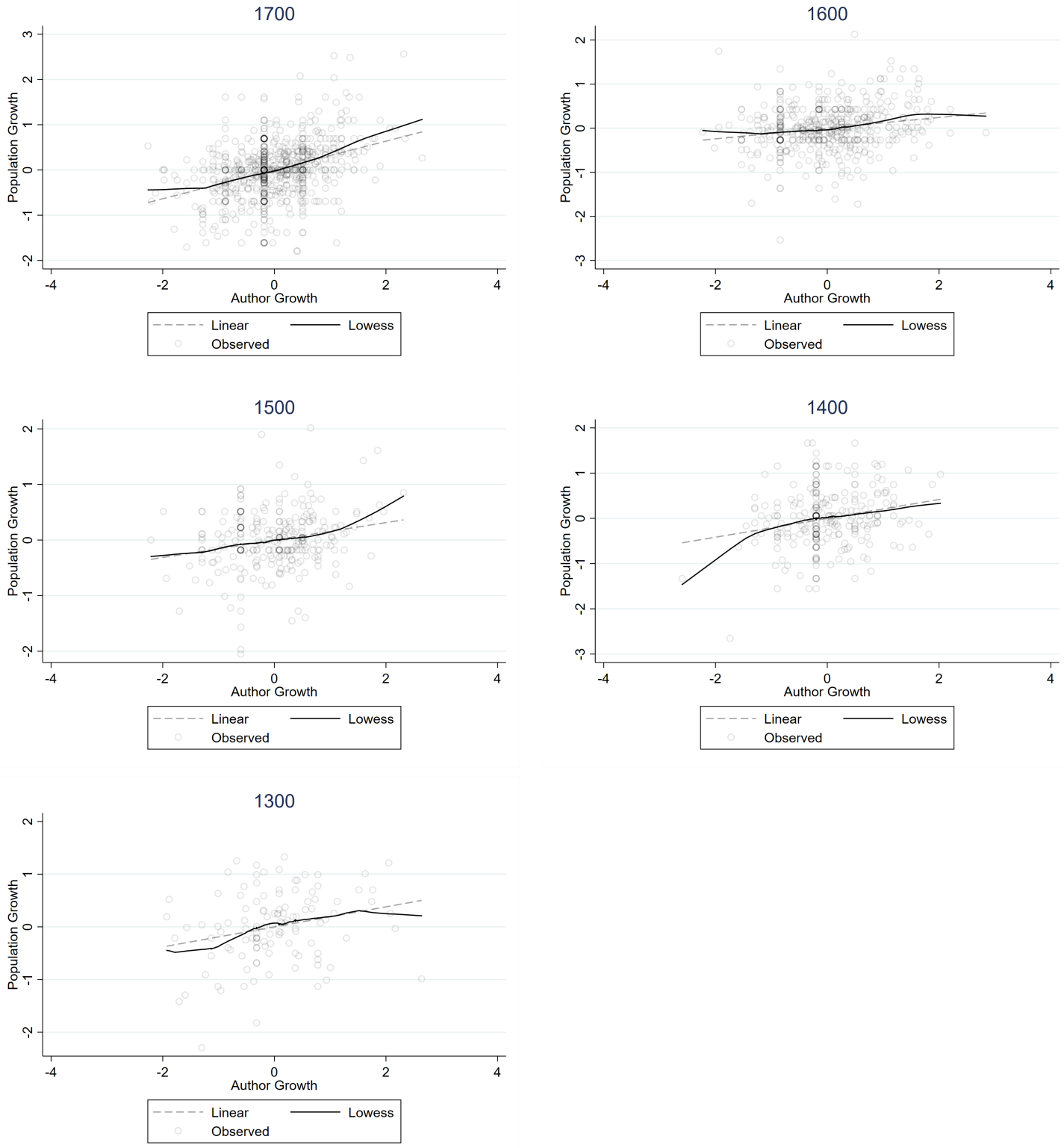

Figure 4: **City and Author Growth**

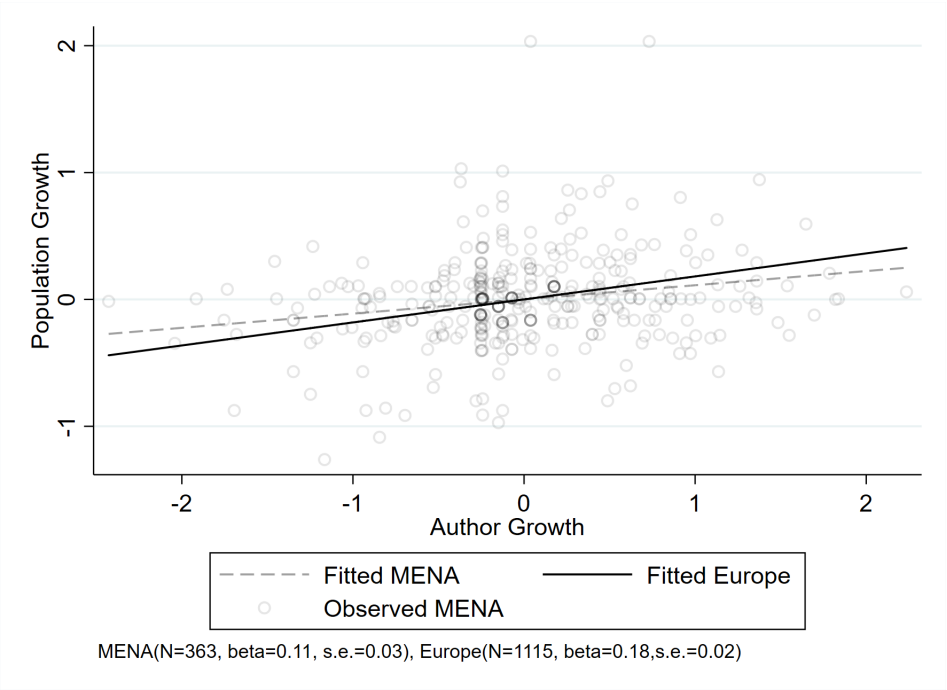

Figure 5: **City and Author Growth: MENA region**

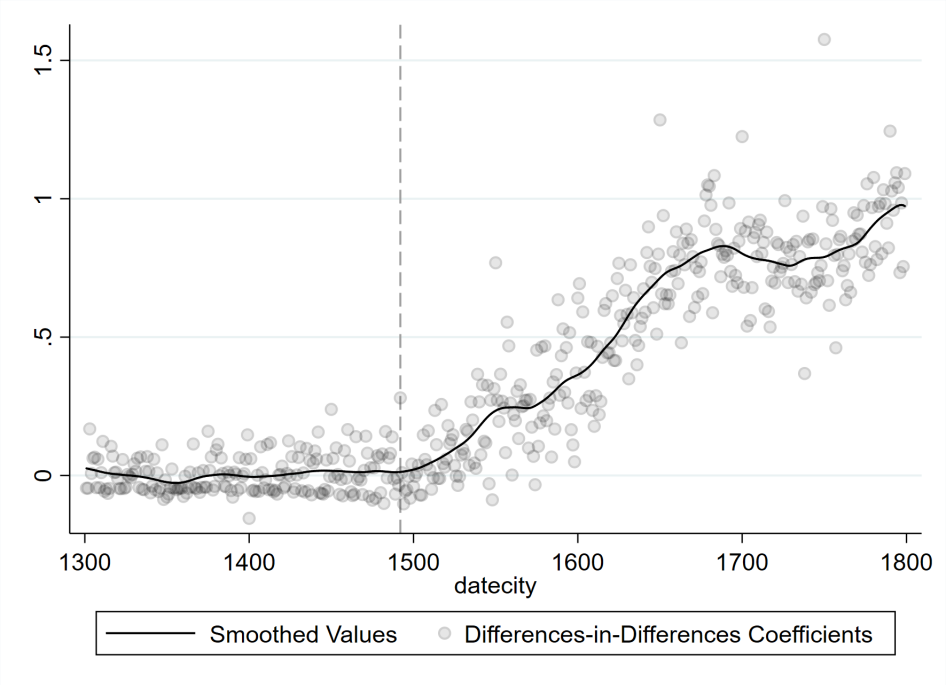

Figure 6: **The Rise of the Atlantic Traders 1300-1800**

Differences-in-Differences Coefficients by Year (omitted year 1425): Vertical line marks 1492