Historical Change in European City Populations: The Emergence of Zipf's Law

Jeremiah Dittmar*

Abstract

Zipf's Law is one of the most robust and widely observed empirical patterns in economics: the distribution of city populations follows a power law. Has this always been the case – or did historical processes governing the distribution and growth of city populations differ from modern ones? This research examines European city populations from the middle ages through the early modern era and documents that Zipf's law has not always held. In the middle ages, Europe's largest cities grew relatively slowly and struggled to become "large enough" to satisfy a power law. In the early modern era, large cities began to grow as fast as smaller cities and Zipf's Law emerged between 1500 and 1800. Historical evidence points to a relaxation of fixed factors that previously constrained city growth – driven by falling trade costs and the expansion of knowledge-based activities and reflecting economic institutions.

Key words: Cities, Growth, Zipf's Law, Population, Economic History, Power Laws

JEL: J1, N9, N34, O1, O4

---

*Contact: Department of Economics, LSE, Houghton Street, London WC2A 2AE, UK. Email: `dittmar.academic@gmail.com`. Tel: +44 (0)20 3486 2881. I thank Brad DeLong, Barry Eichengreen, Xavier Gabaix, Chad Jones, Suresh Naidu, Christina Romer, Noam Yuchtman, and seminar participants at American University, Dartmouth, Stanford, UC Berkeley, UC Davis, University of British Columbia for comments, discussions, and suggestions.

1 Introduction

Zipf's Law is one of the most robust and widely observed empirical patterns in economics: The distribution of city populations follows a power law. A large literature suggests that Zipf's Law is virtually ubiquitous across time and space (Krugman 1996a, Gabaix 1999a) and that the distribution of populations conforming to Zipf's Law may be extremely persistent over many centuries, even in the face of dramatic shocks (Davis and Weinstein 2002). The apparent ubiquity of Zipf's Law suggests more than an underlying order in urban hierarchies. It suggests that fundamental determinants of the distribution of economic activity may be shared by societies with very different social and institutional arrangements. It raises the possibility that the key determinants of the location of economic activity reside in trans-historical fundamentals.

Has Zipf's Law always characterized urban hierarchies – or did historical processes governing the distribution and growth of city populations differ from modern ones? In this research I examine European city population sizes from the middle ages through the early modern era and document that Zipf's law has not always held. In the middle ages, Europe's largest cities grew relatively slowly and struggled to become "large enough" to satisfy a power law. In the early modern era, large cities began to grow as fast as smaller cities and Zipf's Law emerged between 1500 and 1800. Historical evidence points to a relaxation of fixed factors that previously constrained city growth – driven by falling trade costs, the expansion of knowledge-based activities, and economic institutions.

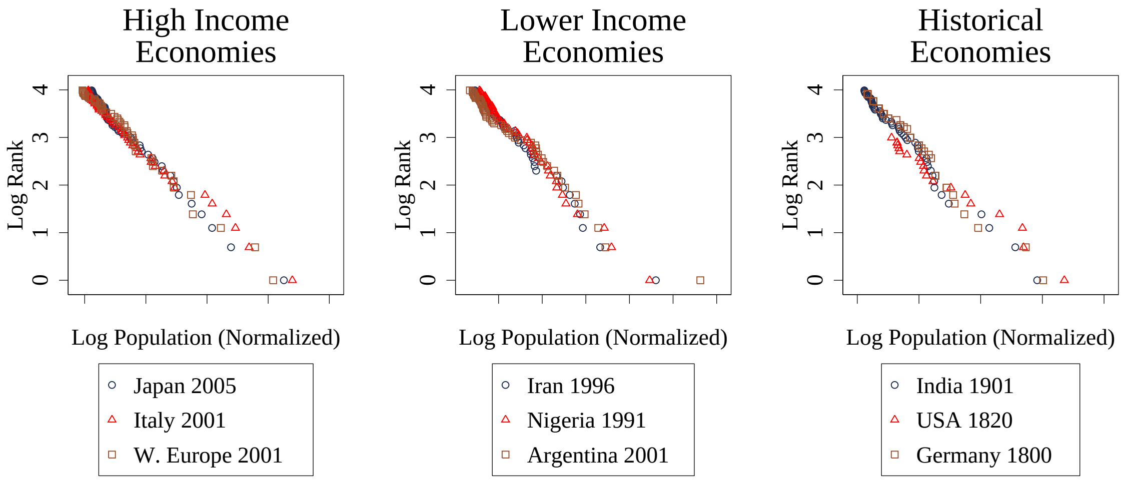
A substantial body of evidence documents the apparent ubiquity of Zipf's Law. Krugman (1996a: 39) observes that the power law pattern governing city populations is so exact and so "suspiciously like a universal law" as to be "spooky." Gabaix (1999a: 129) notes that it appears to hold in all economies and periods for which there are data. The Zipf's Law regularity is most commonly documented graphically, by plotting the *log linear* relationship between city population ranks and city populations that obtains when the urban systems follow a power law.[1] Figure 1 presents evidence that suggests a power law may characterize the distribution of city populations across both higher and lower income economies in the contemporary world – and in the 19th and early 20th centuries. This range of evidence has led economists to wonder whether Zipf's Law describes all urban systems and to develop models of cities which take Zipf's Law as a fundamental stylized fact (Desmet and Henderson 2015; Behrens and Robert-Nicoud 2015).

---

[1] Under Zipf's Law, for cities i with populations S ranked R: $R_i = \alpha S_i^{-\beta}$, with $\beta \approx 1$. Debate exists on whether the contemporary distribution of urban populations follows a power law or is log normal – and, relatedly, over whether the relevant units are administratively defined locations or agglomerations defined according to more clearly economic criteria. These questions are discussed in Section 2 below.

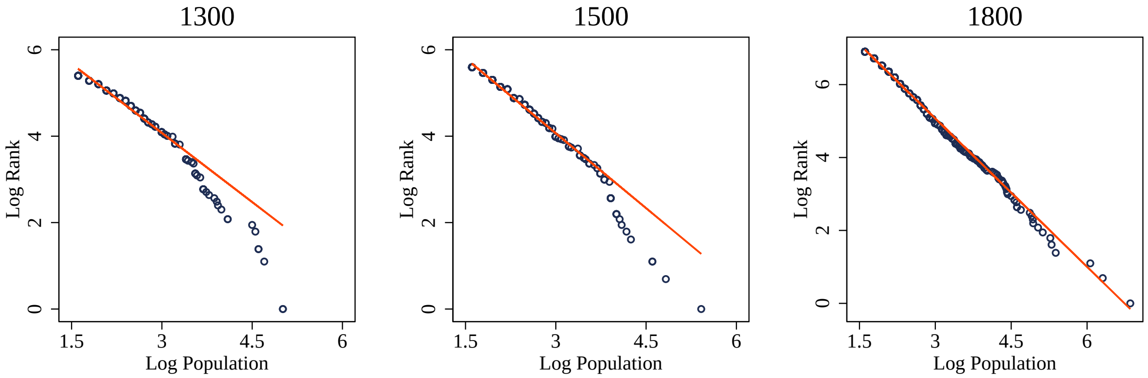
In contrast, this research shows that Zipf's Law only emerged in Western Europe between 1500 and 1800 and studies its emergence as part of the broader economic transition to the era of modern growth. Figure 2 previews the main results in this paper. It shows how the European economy did not exhibit Zipf's Law in 1300 and how Zipf's Law emerged in Western Europe by plotting city population ranks against city populations. In Figure 2, we observe that the largest cities in Western Europe were far smaller than they would have been had they conformed to a power law in 1300 and 1500. Similar deviations from Zipf's Law in the upper tail are observed when we examine historical cities at the regional and proto-national level, as discussed below.

Figure 1: Zipf's Law Across Time and Space



Note: Contemporary data are from Brinkhoff (2008). Historical data for India, Germany, and the USA are from India Office (1905), Bairoch et al. (1988), and Gibson and Jung (2005), respectively.

Figure 2: The Emergence of Zipf's Law in Western Europe



This figure plots raw data on city populations (S_i) and their corresponding size rankings (R_i) and fitted values estimated using robust non-parametric Theil regression and the model: $\ln(R_i) = \alpha - \beta \ln(S_i) + \epsilon_i$. Populations in thousands are from Bairoch, Batou, and Chèvre (1988).

The leading theories tie Zipf's Law to either an underlying distribution of static fundamentals or to random, "size-independent" growth. The most prominent static fundamentals theories emphasize fixed or quasi-fixed geographic advantages – locational fundamentals (Krugman 1996a, Gabaix 2009). Davis and Weinstein (2002: 1269-1270) capture the central thrust of the locational fundamentals view when they observe that, "crucial characteristics for locations have changed little over time…for example, there are advantages of being near a river, on the coast, on a plain instead of a mountain." Recent contributions also model Zipf's Law as a reflection of the distribution of static but non-geographic fundamentals, such as entrepreneurial talent, that may also have a thick upper tail (Behrens, Duranton, and Robert-Nicoud 2014). In contrast, random growth theories observe that power law distributions will emerge when all cities draw growth rates from a common distribution and provide a bench-mark for thinking about city growth.[2] A related body of theory suggests that city size hierarchies are the outcome of a process in which cities grow to optimal size given industry-specific externalities and their industrial specializations (Henderson 1974, Rossi-Hansberg and Wright 2007).

This research uses data on European cities since 800 AD to study the distribution of city populations and to reconsider static fundamentals and random growth theories for Zipf's Law. Prior research has studied non-European economies, different time periods, and evidence on populations of different units, such as counties and regions.[3]

In the data on European cities, we observe that Zipf's Law did not hold before 1500 and emerged by 1800. This fact raises the possibility that Zipf's Law may be explained by locational advantages that are activated by historically specific technologies, institutions, and market relationships.[4] In the data, we observe Zipf's Law emerging only after city growth became size-independent, and that after 1500 cities on navigable water played a key role as locations where size was not correlated with growth. More broadly, we observe significant churning in the distribution of European city populations. Taken together, the evidence suggests that geography matters but is not destiny, and that natural advantages may reflect the level of technological development and/or institutional factors (Acemoglu, Johnson, and Robinson 2005; Bleakley and Lin 2012). These findings characterizing the distribution and growth of cities are supported by a rich body of evidence on the historical

---

[2] A power law emerges as the limiting distribution when random growth combines with a friction – such as some lower barrier or floor on city sizes. Absent this friction, the limiting distribution is log normal. See Gabaix (1999a), Eeckhout (2004), and discussion below.

[3] Davis and Weinstein (2002) study populations of Japanese cities in the 20th century and of regions over many centuries. Michaels, Rauch, and Redding (2012) study urbanization in the USA since 1880 across rural and urban counties. Desmet and Rappaport (2015) US counties since 1800. Ioannides and Zhang (2015) study the size distribution of city walls in historic Chinese cities. Comparisons between this and previous work, including the economic history literature, are discussed further below.

[4] While locational advantages may also constructed – harbors may be dredged and canals dug. Works of this sort were very limited pre-1800, making geographic fundamentals more salient ceteris paribus.

demography of European cities. Moreover, I document that the observed deviations from Zipf's Law cannot be accounted for by measurement error in historical data.

Historical evidence provides guidance for an analytic narrative on how Zipf's Law emerged as the pattern of city growth was transformed in Europe. The deviations from Zipf's Law observed in the middle ages reflected impediments to trade, the operation of markets, and the accumulation of upper tail knowledge. Low agricultural productivity and high trade costs limited the growth of the largest cities. During the early modern era, developments in trade technologies and the emergence of new trade routes combined with advances in agricultural productivity and knowledge-based innovations in urban activities. These changes in the economy transformed city growth and the urban structure of Europe: it became possible for large cities to grow as fast as small cities. Zipf's Law emerged with this "modern" pattern of size independent growth 1500-1800.

The pattern of European city growth and the emergence of Zipf's Law also reflected the institutional environment. Medieval Europe was characterized by political fragmentation and legal restrictions on labor mobility. Political fragmentation impacted cities' ability to feed themselves by raising the costs and risks of food insecurity. The development of a large scale international grain trade and advances in agricultural productivity in early modern Europe relaxed the food constraint on city growth. The institutional dimensions of the challenge facing medieval and early modern cities are highlighted by the fact that different arrangements had supported the provisioning of large cities during the Roman era.[5] The salience of institutions is also underlined by a comparison between Western and Eastern Europe, where legal restrictions on labor mobility were established during the early 1500s. In Western Europe, restrictions on the mobility of tenant farmers were relaxed and rural-to-urban migration became a fundamental determinant of early modern city growth. In Eastern Europe, legal restrictions on mobility were associated with persistent deviations from random growth and Zipf's Law.

This paper thus documents how a modern city system and growth regime emerged in European history – with the development of markets, technology, and institutions. These changes occurred in relatively advanced economies in the centuries preceding the onset of modern, capitalist economic growth which shows up in macro aggregates. Zipf's Law notably developed before the industrial revolution, in an era when cities did not exhibit the industrial or service sector specialization observed in cities today (Nicholas 2003).

---

[5] Over the period we study, the largest cities in Europe grew to have as many inhabitants as ancient Rome. Ancient Rome was supported by an institutional infrastructure that secured large scale grain imports from North Africa and other provinces. In ancient Rome, policy makers introduced food subsidies for residents of the capital and tax incentives for grain merchants. Historical research typically suggests that the population of Rome was about 1 million around year 100 CE. Lower bound estimates suggest populations of at least 400,000. Evidence from Chinese history is discussed below.

2 Literature

2.1 Zipf's Law

Zipf's Law for cities can be characterized in two ways.[6] The first is in terms of the probability distribution of city populations in the upper tail. Where Zipf's Law holds, city populations are distributed according to a power law such that the probability of drawing a city with population size S greater than some threshold N is:

$$\Pr(S > N) = \alpha N^{-\beta} \tag{1}$$

Equation (1) is consistent with a power law distribution where the size ranking of a city (denoted R) is inversely proportional to its population size[7]:

$$R = \alpha S^{-\beta} \tag{2}$$

Equation (2) implies a second characterization of Zipf's Law:

$$\log R = \log \alpha - \beta \log S \tag{3}$$

Zipf's Law is often illustrated by plotting city rank (R) against city size (S). In some cases, the literature associates Zipf's Law with the case where $\beta \cong 1$. However, estimates of β vary across time and economies. This paper focuses on the log-linear (power law type) relationship, but is agnostic on the range of acceptable β's.[8]

2.2 Theories

Two principal types of theories have been advanced to explain Zipf's Law. First, static fundamentals theories explain Zipf's Law as a reflection of static underlying distributions of productivity. Static geographic theories explain Zipf's Law as reflecting the fixed distribution of natural advantages across locations. Static but non-geographic theories

---

[6] The proper entities are urban agglomerations, which are what the data in this research document and what this paper analyzes (see below for description of data). For discussion of how the definition of cities matters, and reasons to use economic as opposed to administrative definitions of agglomerations, in contemporary settings see Rozenfeld, Rybski, Gabaix, and Makse (2011).

[7] Even if the data generating process conforms to equation (1), equation (2) only holds approximately. Gabaix (1999b, 2009) provides discussion and derivations.

[8] Gabaix and Ioannides (2004: 2350) observe: "the debate on Zipf's Law should be cast in terms of how well, or poorly, it fits, rather than whether it can be rejected or not...if the empirical research establishes that the data are well described by a power law with exponent $\beta \in [0.8, 1.2]$, then this is a useful result." NB: For consistency, notation changed to β.

explain Zipf's Law as an equilibrium outcome given a fixed and heavy tailed distribution of talent, agglomeration economies, and congestion costs. Second, random growth theories explain Zipf's Law as the outcome of a growth process in which all cities draw growth rates from some common distribution.

Static fundamentals theories have suggested that city size distributions reflect either the distribution of geographic advantages or of entrepreneurial talent. Krugman (1996b) and Davis and Weinstein (2002) observe that the distribution of propitious locations may follow a power law and thus account for the size distribution of cities.[9] Behrens, Duranton, and Robert-Nicoud (2014) develop a static model of cities in which the talent of entrepreneurs follows an (approximate) power law and is a key determinant of city size. In their model, the fixed distribution of entrepreneurial talent, agglomeration economies, and congestion costs together deliver Zipf's Law. Lee and Li (2013) observe that even if no one fundamental follows a power law or log normal distribution, population may reflect an underlying composite power law of advantage emerging from random variations across multiple dimensions of fundamentals.

Random growth theories have been developed along several lines. Gabaix (1999b, 2009) has shown that Zipf's Law may emerge as the limiting distribution of a process in which cities draw random growth rates from a common distribution, provided there is an arbitrarily small reflecting barrier that prevents cities from getting "too small." This assumption is consistent with the evidence from European history (Livi-Bacci 1999).[10] Recent theoretical work has further explored how random growth may deliver Zipf's Law. Cordoba (2008) provides a model in which either tastes or technologies follow a reflected Brownian motion. Rossi-Hansberg and Wright (2007) develop a model in which there are increasing returns at the local level and constant returns in the aggregate, and Zipf's Law emerges under special circumstances.[11] In Cordoba (2008) and Rossi-Hansberg and Wright (2007), cities specialize in particular final (or tradable) goods, and Zipf's Law emerges as cities reach efficient size given their specialization.

---

[9] As discussed on p. 15, the fact that Zipf's Law emerged over time, and that there was substantial "churning" in Europe's urban hierarchies, indicates that a purely geographic theory will be insufficient. It also suggests that pre-modern growth was non-random.

[10] Livi-Bacci (1999) observes that while certain cities have experienced relative decline, since 1000 AD cities have rarely disappeared in European history. Without this assumption, random growth delivers a lognormal distribution, not a power law. Earlier contributions tying Zipf's Law to random growth include Krugman (1996a) and Simon (1955). See Gabaix (1999b, 2009).

[11] It emerges when (i) capital does not enter production and permanent productivity shocks are the only shocks, or (ii) production is linear in capital and shocks are transitory.

2.3 Existing Evidence

The economic history literature has examined Zipf's Law in a number of settings, but to my knowledge has not examined its emergence in Western Europe.[12] Russell (1972) provides data that can be used to show that the largest cities in the local urban systems of medieval Europe were too small to satisfy Zipf's Law, consistent with the European-wide pattern emphasized in this paper.[13] Stabel (2008) provides data that can be used to document similar deviations from Zipf's Law in the Low Countries in 1450 at both the aggregate and provincial level. Archaelogical data confirm departures from Zipf's Law across a range of pre-modern or non-capitalist economies (Johnson 1980, Drennan and Peterson 2004, and Savage 1997).[14] Thus de Vries (1990: 52) observes that rank-size distributions, "can summarize effectively the process of urbanization and identify gross differences in the design of urban systems over time [and] in different societies."[15]

A broader debate exists in economics over the origins of Zipf's Law. Davis and Weinstein (2002) find that in Japan a *regional* analogue to Zipf's Law held stretching back thousands of years and that the hierarchy of regional population densities in Japan has been relatively stable over many centuries. Davis and Weinstein (2002) also observe that the Japanese *city* size hierarchy has been stable even in the face of massive shocks due to the firebombing of select Japanese cities during the World War II. Based on these findings they argue that fixed locational fundamentals are key determinants of the distribution of populations. A related literature has examined local growth and the existence of Zipf's Law using evidence from North American starting in the 1800s. For example, Desmet and Rappaport (2013) show that US *counties* only gradually converged towards size-independent growth in the 20th century, whereas Zipf's Law appears to characterize city populations in the USA from the early 1800s. Michaels, Rauch, and Redding (2012) document deviations from size-independent growth across sub-county

---

[12] See, for instance, Guérin-Pace (1995), Bairoch (1988), and de Vries (1984).

[13] Russell (1972) partitions urban Europe into sub-regions and examines a central Italian region around Florence, a Southern Italian region around Palermo, a Southern German region around Augsburg, etc.

[14] Zipf's Law has also been examined by anthropologists. Smith (1982) observes that pre-capitalist economies typically do not exhibit Zipf's Law. Smith suggests deviations from Zipf's Law may be due to limited "commercial interchange" or to low agricultural productivity, but does not identify the negative correlation between size and growth as the key source of historical deviations from Zipf's Law.

[15] Deviations from Zipf's Law are observed in economies as varied as 19th century China, Ottoman Palestine, 13th century France, and medieval Tuscany. Interesting recent work by Ioannides and Zhang (2015) finds that area contained inside the historic walls of Chinese cities closely approximate Zipf's Law. Because city populations fluctuated dramatically within given or similar-sized walls this raises an interesting question as to whether China was different, or whether walls and populations are different units. For example, in 1500 almost all of London's population lived with the Roman and medieval city walls (Rowley 2006), with adjustment on the density margin. In constrast, the New City of Kaifeng in China was built with 27 km of walls in 955 CE, but already in 1021 CE an additional 14 large wards outside the walls were formally recognized by city authorities (Morris 2013: 157).

jurisdictions in the USA that include rural locations 1880-2000. However, Ioannides and Overman (2004) show that contemporary city growth in the USA appears to be random. This literature naturally raises the question of whether the evidence concerning the relationship between Zipf's Law and growth processes differs depending on the units we examine or across geographic and historic settings.

A range of evidence exists on the existence and parameterization of Zipf's Law. Soo (2005) examines cross-country data and finds that they are inconsistent with a $\beta = 1$ Zipf's Law in many economies, a finding also emphasised in Ioannides et al. (2008). A debate also exists around the question of whether the entire size distribution of agglomeration populations is Pareto. Eeckhout (2004) documents that in the contemporary USA, the entire size distribution of administratively defined places is lognormal and not Pareto. However, Rozenfeld, Rybski, Gabaix, and Makse (2011) find that when cities (and their populations) are defined according to economic rather than legal or administrative criteria, Zipf's Law holds for cities with populations as small as 12,000 in the USA and 5,000 in the UK. Rossi-Hansberg and Wright (2007) observe that contemporary data are marked by a mild case of what this paper shows was a glaring historical fact: from the perspective of Zipf's Law, small cities are under-represented and big cities are too small. Rossi-Hansberg and Wright argue that this results when small cities grow quickly and large cities grow slowly. Gabaix (1999b) observes a further anomaly: capital cities frequently do not conform to Zipf's Law. I return to these points below.

3 Data

This section presents the city population data and the regional classification of cities. Additional data are discussed as introduced and in Appendix A.

3.1 Data on City Populations

This research employs data on European city populations from Bairoch, Batou, and Chèvre (1988), who record the populations of urban agglomerations (in thousands).[16] The "Bairoch data" comprise cities with at least 5,000 inhabitants. This research focuses on the period from 1300 forward, when data on a relatively large set of cities are available. Table 1 summarizes the Bairoch data for Western Europe. Figure 3 shows the locations

---

[16] Bairoch, Batou, and Chèvre (1988: 289) make a special effort to include, "the 'fauborgs', the 'suburbs', 'communes', 'hamlets', 'quarters', etc. that are directly adjacent" to historic city centers.

of the Western European cities examined in this research.[17]

Table 1: Summary Statistics on City Populations and City Growth

	Population at Start of Period			Annualized Growth Rate		
Time Period	Cities	Mean	St. Dev	Cities	Mean	St. Dev
800 - 900	31	23.9	27.5	9	0.09%	0.52%
900 - 1000	13	27.7	11.3	12	0.30%	0.77%
1000 - 1200	74	29.2	64.0	49	0.16%	0.30%
1200 - 1300	99	22.4	20.8	93	0.14%	0.58%
1300 - 1400	255	17.4	20.0	160	-0.22%	0.58%
1400 - 1500	187	18.6	26.7	155	0.06%	0.52%
1500 - 1600	321	15.4	19.3	285	0.18%	0.46%
1600 - 1700	514	15.5	25.2	456	-0.13%	0.55%
1700 - 1750	539	17.2	39.7	480	0.28%	0.60%
1750 - 1800	686	16.9	40.6	675	0.29%	0.63%
1800 - 1850	1,311	14.2	36.6	1,263	0.68%	0.78%

This table records populations and growth rates for cities with population of 5,000 or more at the beginning of each period. Populations are in thousands. Population growth rates and standard deviations are computed on an annualized basis for cities observed at the beginning and the end of each period.

3.2 Regional Classification of Cities

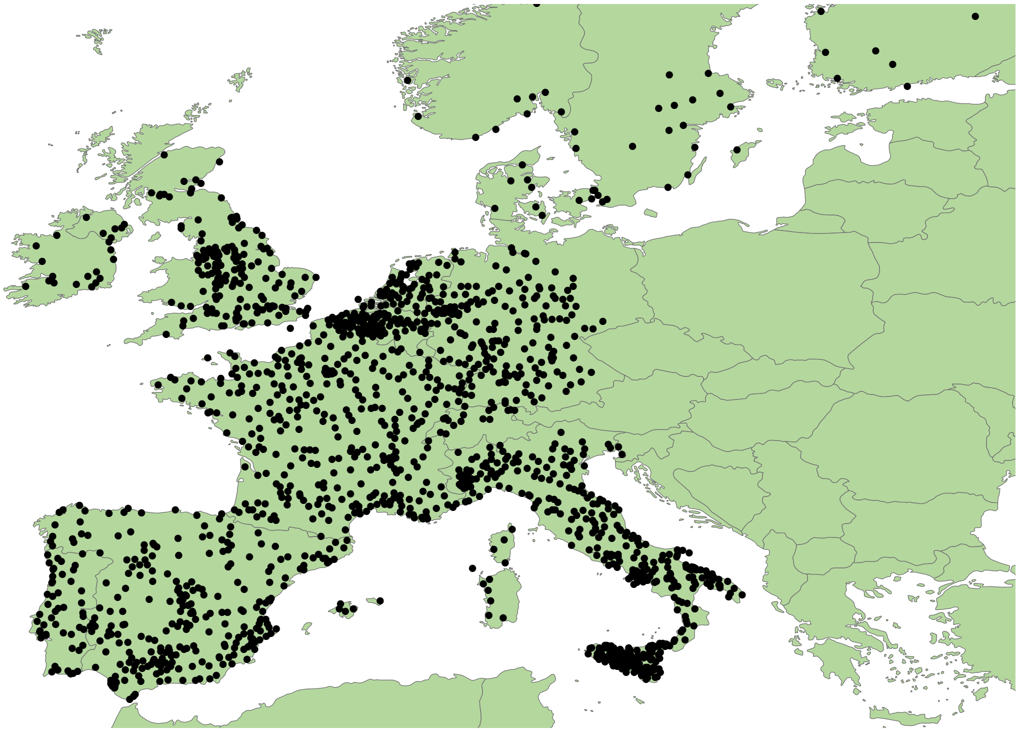
The analysis in this paper examines city populations across Western Europe. This reflects (i) the transnational environment and exchanges characterizing city growth in early modern Europe and (ii) differences in the institutional environment that distinguished cities in Western Europe from cities in Eastern and Ottoman Europe.

A transnational perspective is appropriate for an analysis of city growth in European history. Between 800 and 1800, political fragmentation allowed cross-border economic linkages to organize urbanization and for European cities to begin to develop a single, integrated urban system (de Vries 1984; Nicholas 2003; Jones 1981; and Bosker, Buringh, and van Zanden 2008). Economic integration in Europe in part reflected the institutional environment, including the diffusion of Roman and merchant law in the middle ages, which supported exchange and mercantile activities that were concentrated in cities (Berman 1983; Berman and Kaufman 1978), as well as the self-enforcing institution of the

---

[17] For select cities Bairoch, Batou, and Chèvre (1988) observe populations from periods before cities reach 5,000 inhabitants. This research examines cities with population of at least 5,000 to address concerns about how the non-random selection of fast growing cities into the sample would shift the distribution of growth at the lower end of the distribution. Note also that the results are robust to using 10,000 or 20,000 as the population cut off for examination, thresholds around which there is no selection into observation.

Figure 3: Historic cities of Western Europe



community responsibility system (Greif 2006). Significantly, the deviations from Zipf's Law documented in this paper are not figments of the aggregation. The emergence of Zipf's Law in Western Europe between 1500 and 1800 was mirrored by its emergence over the same period at the local and national level, as documented below.

The distinction between Western and non-Western cities was determined by differences in institutions providing municipal autonomy for cities and governing mobility between the rural and urban sectors. Cities West of the Elbe River, which cuts through Eastern Germany, developed in an environment where town charters guaranteed the right to legal proceedings in town courts, the right to sell homes and to move, and freedom from obligations of serfdom. These institutions fostered geographic mobility, relatively secure property rights, and urban commerce.[18] The institutional environment was different in Eastern and Ottoman-controlled Europe. In Eastern Europe, legal institutions limiting labor mobility and city autonomy were installed after 1500. These laws tied tenant farmers to rural estates, provided for the return of fugitive serfs, and limited the activities of urban merchants. An extensive literature documents the importance of the Elbe River as an institutional boundary distinguishing Western Europe from territories in

---

[18] See Pirenne (1927), Braudel (1979a, 1979b), Friedrichs (1995), Nicholas (2003), Scott (2005), Bideleux and Jeffries (2007), and Bosker, Buringh, and van Zanden et al. (2008).

which serfdom was reinforced after 1500.[19] Under the Ottomans, cities were not granted municipal autonomy, allocations were more heavily influenced by administrative means, and city growth was shaped by the "ruralization" of Christian populations.[20]

This paper therefore focuses on the population distribution and growth dynamics of Western European cities located West of the Elbe River and/or its tributary the Saale and outside Ottoman Europe.[21] I discuss cities in Eastern Europe in Section 6 below.

4 The Medieval Absence and Early Modern Emergence of Zipf's Law

4.1 The Distribution of City Populations

The fact that Zipf's Law emerged over time can be documented graphically and using OLS, quantile, and robust regression.

Figure 4 provides a more detailed view of the evolution of city size distributions between 1300 and 1800 in Western Europe. It shows that prior to 1600 the large cities were "too small," and how Zipf's Law emerged over time, by plotting observed populations against fitted values associated with the robust non-parametric regression estimator proposed by Theil (1950).[22]

Table 2 provides quantitative evidence that deviations from Zipf's Law went from being large in 1300 to small in 1800. Table 2 measures the historical deviations from the hypothetical Zipf's Law fitted in Figure 4.

Quantile regression can be used to document more precisely where over the range of city sizes the curvature in the rank-size relation emerges.[23] Table 3 presents historical

---

[19] See Kriedte (1979), Berend (1986), Robisheaux (1998), Bideleux and Jeffries (2007), Süchs (1988), Maddalena (1977), Brenner (1974), Anderson (1974a), and Blum (1957).

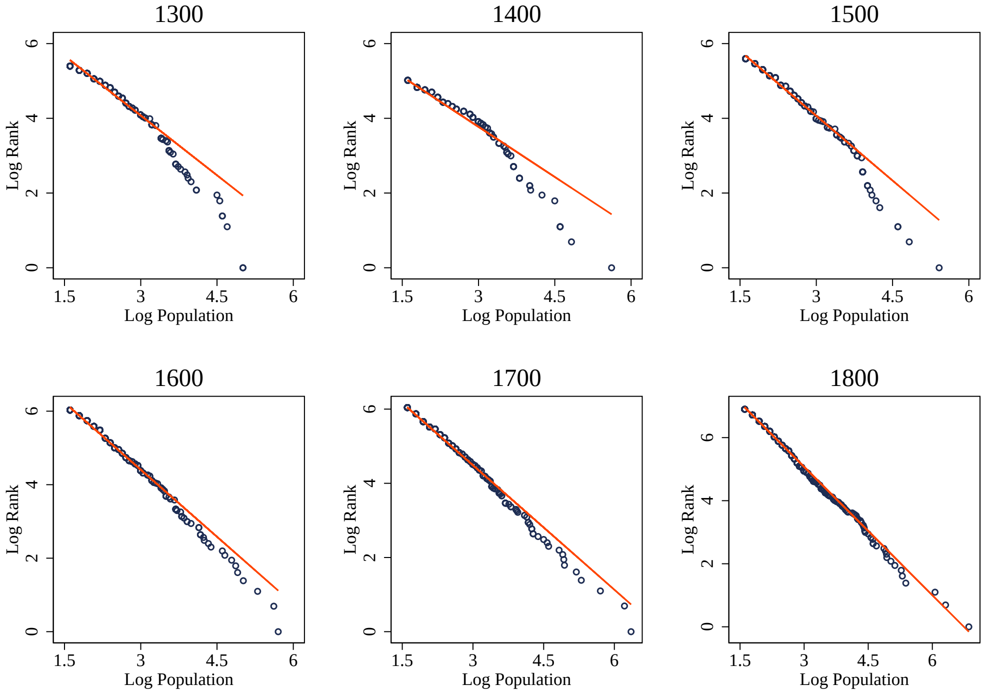
[20] See Stoianovich (1994), Sugar (1977), Bairoch (1988), Hohenberg and Lees (1985), de Vries (1984), Bideleux and Jeffries (2007), and Anderson (1974b).

[21] Cities in Western Germany, the Netherlands, Italy, Belgium, Iberia, Switzerland, and England.

[22] For estimating power law exponents, the Theil estimator is superior to OLS and competitive with the adjusted-OLS estimator proposed by Gabaix and Ibragimov (2011) in its small sample properties and superior in precision, as documented in Appendix C. As a robust regression, the Theil estimator has an advantage over OLS and Gabaix-Ibragimov if outliers may not follow the power law distribution. The way robust regression can be used to gauge departures from power laws is discussed below.

[23] Quantile regression relaxes an assumption the OLS estimator embodies: that, given independent covariates, conditional quantile functions of the response variable have a common slope. Quantile regression estimates assume a piecewise linear loss function and minimizing the (asymmetric except in the case where $\tau = 0.5$) sum of absolute residuals. See Koenker (2005).

Figure 4: The Emergence of Zipf's Law in Western Europe



This figure plots (1) raw data on city populations (S_i) and their corresponding size rankings (R_i), and (2) fitted values estimated using robust non-parametric Theil regression and the model: $\ln(R_i) = \alpha - \beta \ln(S_i) + \epsilon_i$. Populations in thousands are from Bairoch, Batou, and Chèvre (1988).

Table 2: Mean Square Deviations from Estimated Power Law

Year	Mean Square Deviation
1300	6.27%
1400	4.21%
1500	1.51%
1600	0.58%
1700	0.50%
1800	0.18%

This table presents the mean square deviation from power law estimated with the Theil regression estimator. For cities indexed with $i = 1, \ldots, N$, actual (observed) population S_i^a, and Zipf-consistent population S_i^z computed from Theil regression estimates, mean square deviation is: $MSD = N^{-1} \sum_{i=1}^{N} (S_i^a / S_i^z - 1)^2$.

estimates of local, quantile slope parameters associated with equation (3). It shows that the big non-linearities were at the upper end of the city size distributions. In Table 3, as τ declines the estimates describe the local Zipf exponents (slopes) associated with progressively larger cities.[24] That the big non-linearities are located at the upper end of the city size distribution is evident in the fact that local slopes change modestly as τ falls

---

[24] The parameter τ defines quantiles in the response variable, city rank. The τ quantile in the city rank distribution corresponds to the $(1 - \tau)$ quantile in the city size distribution.

from 0.75 to 0.25 and sharply as τ falls from 0.25 to 0.1. By 1800 the local Zipf exponents of Western European cities are relatively stable in the upper tail (as τ declines) and fall within the (0.7, 1.5) range observed in contemporary economies (Soo 2005).

A formal test rejects the null hypothesis that the data follow a power law distribution up through 1500. Indexing cities with i and denoting city size S and city rank R, the test developed in Gabaix (2009) relies on an OLS regression:

$$\ln(R_i - 1/2) = \delta_0 + \delta_1 \ln S_i + \delta_2(\ln S_i - S^*)^2 + \epsilon_i \tag{4}$$

where $S^* \equiv \text{cov}[(\ln S_i)^2, \ln S_i]/2\text{var}[\ln S_i]$ and the shift of -1/2 provides the optimal reduction in small sample bias in the OLS setting.[25] Under the Gabaix test, we reject the null hypothesis of a power law with 95 percent confidence if and only if $|\hat{\delta_2}/\hat{\delta_1}^2| > 1.95(2n)^{-0.5}$. Table 3 (column 7) shows that implementing the Gabaix (2009) test that we can reject Zipf's Law in Western Europe through 1600, but not from 1700 forwards. (Complete results and underlying estimates are presented in the Appendix.)

Table 3: Quantile Estimates of Zipf Exponent and Power Law Test

(1)	(2)	(3)	(4)	(5)	(6)	(7)
		Quantile Slope Parameters				Gabaix
Year	Cities	$\tau = 75\%$	$\tau = 50\%$	$\tau = 25\%$	$\tau = 10\%$	Test
1400	187	-0.81***	-0.90***	-1.11***	-1.22***	Reject PL
		(0.01)	(0.03)	(0.08)	(0.04)	
1500	321	-1.16***	-1.19***	-1.16***	-1.42***	Reject PL
		(0.01)	(0.01)	(0.01)	(0.06)	
1600	514	-1.27***	-1.24***	-1.24***	-1.32***	Reject PL
		(0.01)	(0.01)	(0.03)	(0.01)	
1700	539	-1.12***	-1.13***	-1.19***	-1.24***	
		(0.00)	(0.01)	(0.01)	(0.01)	
1800	1,311	-1.37***	-1.39***	-1.39***	-1.41***	
		(0.02)	(0.01)	(0.00)	(0.00)	

This table presents quantile slope parameter estimates and the Gabaix (2009) test for power law distributions. The quantile slope parameter $\beta(\tau)$ is estimated with regression: $\ln R_i = \alpha - \beta(\tau)\ln S_i + \epsilon_i$. As τ declines, quantile regression estimates describe the local slope associated with progressively larger cities (column 6 estimates the slope for at the top 10% quantile). Bootstrapped standard errors in parentheses. Column 7 reports results of the Gabaix (2009) test for power law distributions: "Reject PL" indicates rejection at 95% confidence. See text for description of the text. Full estimates underlying the Gabaix test are reported in the Appendix.

---

[25] An earlier literature examined Zipf's Law with regressions: $\ln(R_i) = \beta_0 + \beta_1 \ln S_i + \beta_2 \ln S_i^2 + \nu_i$. As discussed in Gabaix (2009), heteroskedasticity-robust standard errors will be biased down in this specification and the statistical significance of $\hat{\beta}_2$ is not a robust criterion for a test of Zipf's Law. However, to facilitate comparison with existing studies, Appendix D presents results from this specification which support the conclusion that Zipf's Law emerged in Western Europe 1500-1800.

4.2 Implications for Theory

The observed deviations from Zipf's Law have three key implications for economic theory.

First, the fact that Zipf's Law emerged over time suggests that time invariant factors cannot by themselves explain its existence. The principal geographic features of the European landscape – such as the location of navigable rivers and of bays suitable for ports – remained essentially unaltered before 1800. The fact that Zipf's Law emerged between 1500 and 1800 therefore suggests that it is due to something beyond a time invariant power law distribution of propitious locations, or even a time invariant random distribution of advantages along multiple dimensions as in Lee and Li (2013). The fact that Zipf's Law emerged over time is similarly problematic for static models that propose an equilibrium distribution of city populations due to the interactions between fixed distributions of talent, agglomeration economies, and congestion costs. If geography or talent explain Zipf's Law, time varying factors must have changed the returns to geography or talent – or transformed the upper tail distribution of talent.

Second, something other than specialization in goods production accounts for Zipf's Law. Many models of urban hierarchies assume industrial specialization accounts for city size distributions (Henderson 1974; Black and Henderson 1999; Cordoba 2008; Rossi-Hansberg and Wright 2007). In these models, industry-specific externalities combine with diseconomies that increase in city size, driving cities to specialize in specific tradable industries and to optimal size given their activities. But Figure 4 shows that a "modern" urban hierarchy emerged before the widespread adoption of the factory system, when industrial specialization, inter-city trade, and even the non-industrial functional specialization of cities were relatively limited. Nicholas (2003: 7) observes: "Probably no pre-modern city was as functionally specialised as modern industrial cities tend to be."

Third, if random growth is *the* explanation for the rank-size regularity, the fact that this regularity emerged relatively recently implies that there was persistent non-randomness in urban growth in the pre-modern era and leads to the testable hypothesis that random growth emerged in our period.

4.3 Evidence on City Growth over Time

The leading dynamic theories that account for Zipf's Law posit random growth. This section first establishes when and how random growth emerged in Western Europe. It then examines heterogeneity in the emergence of random growth across (i) cities more and less bound by a local land constraint and (ii) young and old cities.

The nature and evolution of pre-modern city growth can be seen when we group cities into size quantiles and examine the distribution of growth rates across quantiles. Figure 5 shows that large cities were at a pronounced growth disadvantage 800-1200, and how this disadvantage declined, by presenting box-plots of growth by population quintile: quintile 1 comprises the smallest cities and quintile 5 the largest. Figure 5 shows that random growth emerged from 1500 forward.

Figure 6 compares the distribution of growth rates for the top 10 percent and bottom 90 percent of cities using nonparametric kernel densities. It shows that through the period 1500-1600, the largest cities consistently grew relatively slowly, and that by 1700-1800 large and small cities were drawing growth rates from similar distributions.

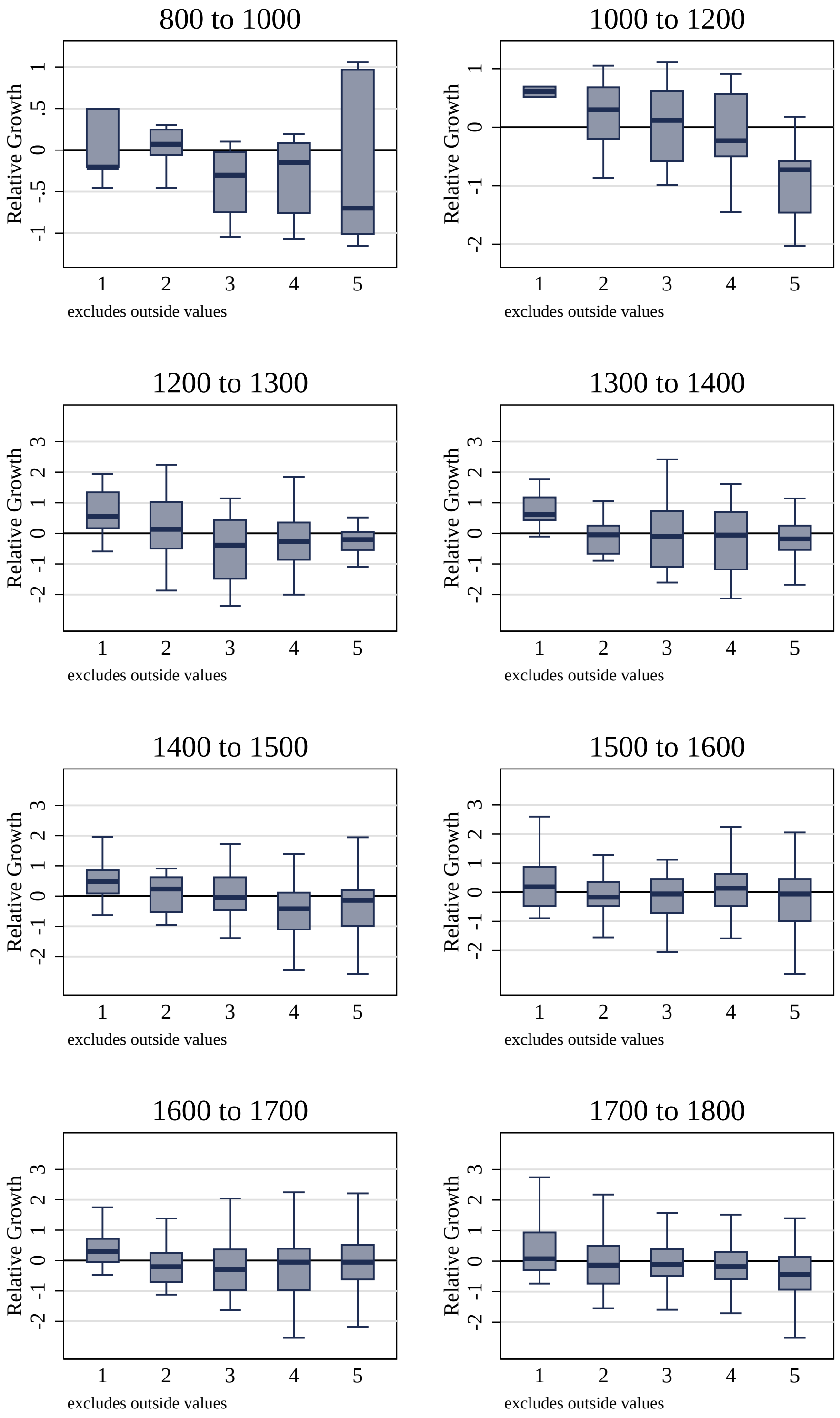
Historical evidence and the existing literature suggest two leading hypotheses about how the relationship between city size and growth may vary across heterogeneous cities. The first hypothesis relates to time-varying returns to, and city growth effects of, geographic advantages. Historical evidence on the growth and provisioning challenges facing large cities (Botero 1602; Pounds 1979; Nicolas 2003) lead us to expect a differential change in the relationship between growth and city size for cities with natural access to "less fixed" land endowments via water borne transport. The historical evidence motivates a first hypothesis: that growth was constrained for all large cities in a world with relatively high transport costs for water borne transport and when the returns to mercantile activities were low, and that cities with access to water borne transport were likely to benefit more after transport costs fell and the returns to trade rose.

Table 6 shows that a similar negative relationship between growth and size existed for all cities through 1500, but that after 1500 we observe sharp differences in growth patterns across cities with access to water transport and landlocked cities. After 1500, there is no significant relationship between growth and population for cities with access to navigable water. For landlocked cities, the significant negative relationship persists.[26]

---

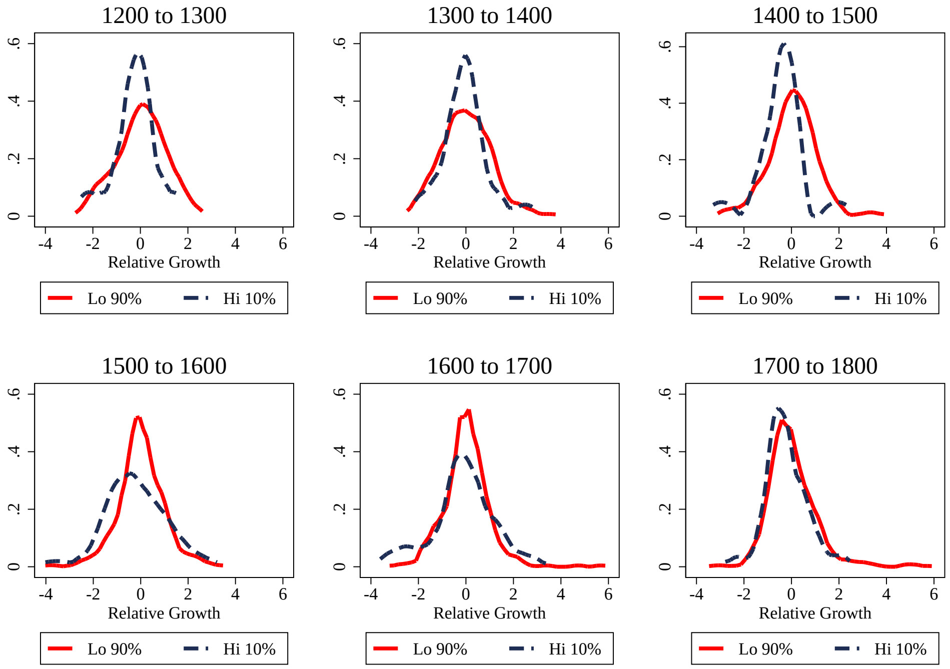
[26] Before 1500, there was a significant negative correlation between size and subsequent growth for all cities (column 2) in every period except 1300-1400. During the 1300s, Europe was struck by the Black Death. The Black Death killed approximately 1/3 of the European population, generating both overall negative city growth and demographic shocks that were highly variable across cities.

Figure 5: The Distribution of City Growth Rates by Size Quantile 800 to 1800



This graph presents the distribution of city growth rates by initial population quintile between 800CE and 1800CE. In each period, the smallest cities are in quintile 1 and the largest are in quintile 5. The boxes describe the interquartile range. The line within each box is that quintile's median. The "whiskers" mark the adjacent values. Growth rates are normalized in standard deviation units. If the population growth rate of city i is g_{it} in period t, and the mean and standard deviation across cities are $\bar{g}_t$ and σ_t, then normalized growth is $\hat{g}_{it} = (g_{it} - \bar{g}_t)/\sigma_t$.

Figure 6: The Distribution of City Growth Rates by City Size



This graph presents the distribution of city growth rates for the largest 10 percent (Hi 10%) and the smallest 90 percent (Lo 90%) of cities. See note to Figure 5 for calculation of normalized growth rates.

Table 4: Correlations Between Growth and Initial Size for Cities with Different Characteristics

		By Transport Access to Navigable Waterways		By Agricultural Potential				By Spatial Proximity to Large Cities	
	All			Old World Crops		Potato Cultivation			
Period	Cities	Navigable	Landlocked	High	Low	High	Low	Close	Far
1000-1200	-0.69**	-0.78**	-0.46**	-0.18	-0.82**	-0.39	-0.81**	.	-0.69**
1200-1300	-0.23**	-0.24	-0.26*	-0.19	-0.24	-0.18	-0.26	.	-0.23**
1300-1400	-0.09	-0.07	-0.14	-0.00	-0.17	0.08	-0.19	-0.06	-0.15
1400-1500	-0.27**	-0.28**	-0.38**	-0.27**	-0.32**	-0.27**	-0.32**	-0.27**	-0.31**
1500-1600	-0.05	-0.05	-0.19**	-0.04	-0.08	-0.06	-0.04	0.00	-0.15
1600-1700	0.00	0.01	-0.18**	0.08	-0.07	0.08	-0.07	0.14	-0.20**
1700-1800	-0.07	-0.08	-0.19**	-0.08	-0.07	-0.06	-0.10	-0.08	-0.09

This table presents correlations between normalized city sizes and growth rates. Cities "On Navigable Water" are located at ocean or sea ports and/or on navigable rivers. "Landlocked" cities are the remaining, non-navigable cities. City agricultural potential for old world crops and for the potato are from Nunn and Qian (2011), and provide a measure of potential agricultural productivity in a 100 kilometer neighborhood of a given city. "High" and "Low" potential cities are those with above and below median potential productivity. Proximity to large cities is measured by an indicator variable for cities with neighbors that (i) were within 20 kilometers and (ii) had 10,000+ inhabitants in a given city-period, following Bosker and Buringh (2017). Significance at the 95 and 90 percent confidence levels denoted "**" and "*", respectively. See note to Figure 5 for calculation of normalized growth rates.

Table 5: Correlations Between City Size and City Growth

	All	Transportation on Water		Proximity to Large Neighbors		Productivity Due to Potato	
Period	Cities	Navigable	Landlocked	Close	Far	High	Low
1000-1200	-0.69**	-0.78**	-0.46**	no	no	-0.39	-0.81**
1200-1300	-0.23**	-0.24	-0.26*	no	no	-0.18	-0.26
1300-1400	-0.09	-0.07	-0.14	no	no	0.08	-0.19
1400-1500	-0.27**	-0.28**	-0.38**	no	no	-0.27**	-0.32**
1500-1600	-0.05	-0.05	-0.19**	no	no	-0.06	-0.04
1600-1700	0.00	0.01	-0.18**	no	no	0.08	-0.07
1700-1800	-0.07	-0.08	-0.19**	no	no	-0.06	-0.10

This table presents correlations between normalized city sizes and growth rates. Cities "On Navigable Water" are located at ocean or sea ports and/or on navigable rivers. "Landlocked" cities are the remaining, non-navigable cities. Significance at the 95 and 90 percent confidence levels denoted "**" and "*", respectively. See note to Figure 5 for calculation of normalized growth rates.

Table 6: Correlations Between City Size and City Growth

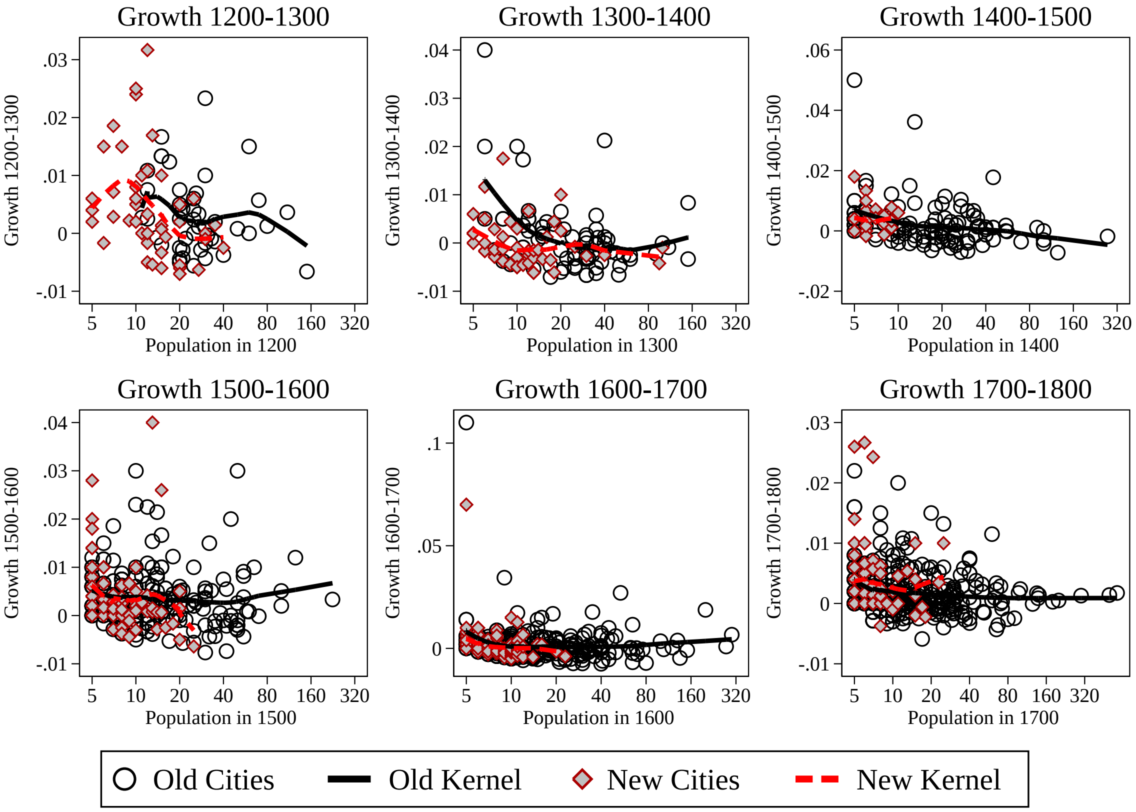
Time Period	All Cities	On Navigable Water	Landlocked
1000 - 1200	-0.69**	-0.78**	-0.46**
1200 - 1300	-0.23**	-0.24	-0.26*
1300 - 1400	-0.09	-0.07	-0.14
1400 - 1500	-0.27**	-0.28**	-0.38**
1500 - 1600	-0.05	-0.05	-0.19**
1600 - 1700	0.00	0.01	-0.18**
1700 - 1800	-0.07	-0.08	-0.19**

This table presents correlations between normalized city sizes and growth rates. Cities "On Navigable Water" are located at ocean or sea ports and/or on navigable rivers. "Landlocked" cities are the remaining, non-navigable cities. Significance at the 95 and 90 percent confidence levels denoted "**" and "*", respectively. See note to Figure 5 for calculation of normalized growth rates.

The second hypothesis relates to potential differences in growth across cities that are young and old. Desmet and Rappaport (2013) find that in US history new ("young") county locations grew systematically faster than old locations in the 1800s and observe that convergence to size independent growth occurred once the extensive margin was closed and the set of locations became more or less fixed. This raises a natural question: Does the appearance of new cities account for the departures from random growth observed in Europe pre-1500? Figure 7 presents scatter plots and kernel regressions of the relationship between city growth and city size for young and old cities, where young cities are those just entering the data in a given century. Figure 7 shows the long run European city growth pattern is different from what Desmet and Rappaport (2013) observe for 19th century US counties. Conditional on size, new and old cities appear to have grown at similar rates in European history, with the possible exception of the period

1500-1600. Before 1500, similar deviations from size independent growth in European data are observed for both old and new cities, indicating that deviations from random growth in Europe were not driven by the fast growth of cities newly entering the urban system.[27]

Figure 7: City Growth Rates and City Population Sizes



This figure presents scatter plots and kernel density regressions of annualized population growth rates on population for "new" and "old". A city is classified as new when it first appears with 5,000+ inhabitants in the Bairoch, Batou, and Chévre (1988) data. Cities are classified as old if they appeared with 5,000+ inhabitants in previous centuries.

It is important to observe that measurement error cannot account for the observed deviations from Zipf's Law. Appendix B provides evidence showing that the population short-falls in the upper tail are so big that they cannot be due to undercounting.[28] For exmple, taking the five largest European cities in 1500: Paris and Naples would have had to have been three times larger than observed to satisfy Zipf's Law, while Venice,

---

[27] These findings hold under alternate definitions of new and old cities. For example, in the subset of German cities we have dates of city establishment and of city incorporation from the *Deutsches Städtebuch*. Among German cities we similarly see no difference in the relationship between size and growth across young and old cities, with city age measured in either of these ways.

[28] As shown in Appendix B, a comparison of Bairoch data to the data in de Vries (1984) reveals no evidence that big city populations are systematically undercounted in the Bairoch data.

Milan, and Grenada would have had to have been more than twice as large as they were. I also show that if missing or mismeasured data for small cities were to account for observed deviations from Zipf's Law this would imply implausibly high urbanization rates 1300-1500. In addition, the observed deviations from Zipf's Law are consistent with a rich evidence on the historical demography of European cities, as discussed in Section 5 below.

5 Historical Evidence on City Populations and Growth

5.1 Constraints on Growth and City Sizes

Historically, the distributions of city growth rates and city populations were determined jointly by agricultural technology, trade technology, returns to activities that were not land intensive, and institutions. This section discusses why effective land was a quasi-fixed factor for pre-modern cities, how this limited the growth of large cities prior to 1500, and how this changed after 1500. It also provides a discussion of the demography of pre- and early modern cities, changes in returns to knowledge-based activities, and institutions.

Historically, transport costs and the risks associated with long distance trade in food constrained cities to rely on local sources for land-intensive wage goods. Contemporaries recognized that this constraint prevented the proportionate growth (size-independent growth rates) associated with Zipf's Law via random growth theory. For example, in 1602 Giovanni Botero observed that, "cities once grown to a greatness increase not onward according to that proportion." Botero observed that the absence of proportionate growth was explained by a fundamental feature of the environment: the difficulty large cities had in feeding themselves given prevailing transport costs (Botero 1602, Book 2, Pt. 9).[29]

Transportation costs were thus a key constraint. Transportation costs were very high, especially for heavy products and overland transport. Grain transported 200 kilometers overland could see its price rise by nearly 100 percent. While the early modern period saw major developments in the international trade in grain, most cities remained heavily reliant on the provision of foodstuffs from a within a circle of 20 to 30 kilometers which avoided heavy transport costs and the risks of reliance on foreign supplies.[30] For this

---

[29] Botero explicitly rejects explanations centered on wars and plagues as giving "no satisfaction." Recent work by Dincecco and Ornato (2014) suggests that local exposure to historical conflict may in fact have increased city growth through a "safe habour" effect where rural and town dwellers sought the relative safety of cities.

[30] See Pounds (1979: 61), Nicholas (2003: 43), and Braudel (1979a: 133). Ballaux and Blondé (2004)

reason, cities preserved forms of land-intensive production. There were gardens, fields, and areas devoted to livestock *within* cities.[31] Costs associated with the transport of fuel generated similar bottlenecks (Ballaux and Blondé 2004).

Agricultural technology and trade costs interacted with political and institutional factors. Because historical agriculture generated limited surpluses, weather shocks could precipitate famine absent secure access to food supplies. Secure access to food supplies was a precondition for the growth of large cities (Maddalena 1977, Pounds 1979) and depended on the economic and political control cities exercised over their hinterlands and surrounding countryside. For Paris, the largest city in 17th century Europe, the problem of securing foodstuffs was especially acute, and is repeatedly stressed by contemporary commentators. In 1591, Pope Gregory XIV issued an edict designed to facilitate the provisioning of Rome from its countryside. In Northern Italy, cities like Milan and Florence conquered and dominated dependent territories that included smaller cities and agricultural hinterlands. Cities on the Istrian and Dalmatian coast similarly controlled territories that stretched inland to the mountains. Venice used political power to compel her colony at Ragusa (Dubrovnik) to export grain to Venice in the 1200s. However, while a city's ability to control a rural district was typically contingent on the absence of a strong regional prince, urban territorial expansion was most often the result of purchases, foreclosed mortgages, and piecemeal treaty acquisitions – and not military conquest. In Germany, Nürnberg, Ulm, and Schwäbish Hall acquired hinterlands of 1,200, 830, and 330 km^2, respectively. The balance of political and economic influence might differ, but similar struggles emerged: Lübeck and Hamburg experienced a series of conflicts with the counts of Schleswig-Holstein and kings of Denmark over rival claims on land, waterways, and resources.[32] Cities that became reliant on foodstuffs imported over distance were exposed to much greater risks when harvests failed (Miskimin 1975).

The risks and costs associated with importing food reflected political fragmentation and the level of state capacity. Ancient Rome had approximately 1 million residents supported by an imperial infrastructure facilitating international grain flows and providing food subsidies for residents of Rome itself.

The fact that land – or a land-intensive intermediate – was a quasi-fixed factor in urban production, is reflected in price data. Kriedte (1979: 27) notes that in the late 16th century grain and oxen prices were 89 and 270 percent higher respectively in Antwerp, commercial hub of the relatively urbanized Low Countries, than in Danzig, the principal

---

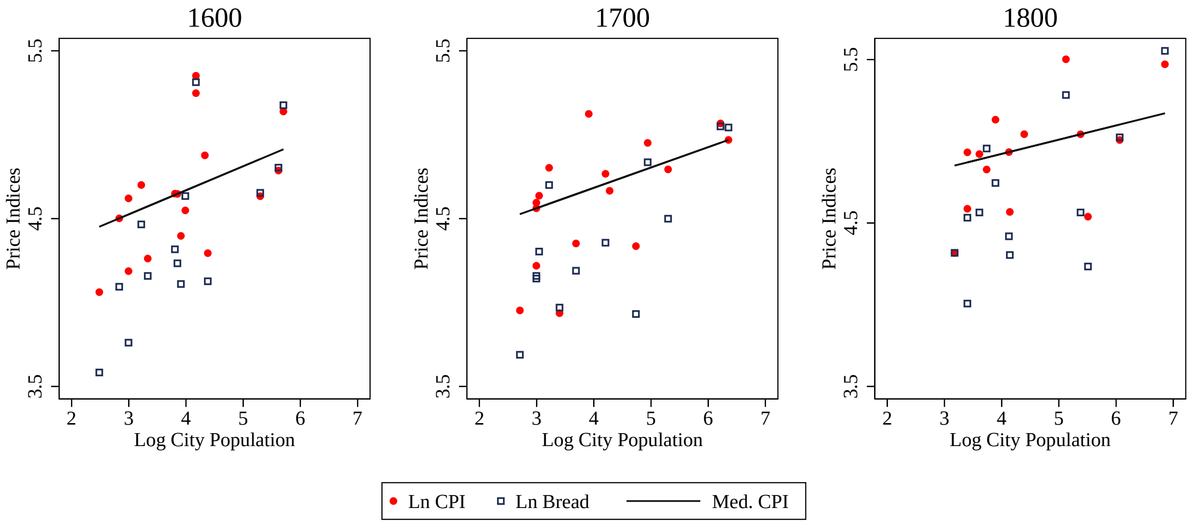
observe transport over land was four times more expensive than on navigable rivers.

[31] Braudel (1979a), Nicholas (2003), Scott (2004), and Friedrichs (1995).

[32] See Scott (2005), de Vries (1976), Pounds (1990), Weber (1958), Pirenne (1958), Chittolini (1994), Livet (2003), Blockmans (1994), Nicholas (2003), Braudel (1979c).

port of rural, grain-exporting Poland. Pounds (1979: 61) similarly observes that prices of agricultural products were increasing in town size.[33] The data support these observations. Figure 8 plots consumer prices and bread prices from Allen (2008) against city population, along with the fitted values from a median regression of consumer prices on city size, and shows that consumer prices tracked bread prices and that prices were correlated with city size.[34] It also provides suggestive evidence that the gradient of price on population declined over time, consistent with the view that while food prices were increasing in city size, the land constraint softened over time.[35]

Figure 8: Consumer Prices and City Populations



This graph plots consumer price index (Ln CPI) and bread price (Ln Bread) data against city populations. The fitted line is the median quantile regression estimate for the CPI. Price data are from Allen (2008). City populations are from Bairoch, Batou, and Chèvre (1988).

5.2 The Relaxation of Constraints on City Growth and Size

Several factors relaxed the land constraint on city growth in Western Europe starting around 1500. First, there were substantial transport cost reductions and increased

---

[33] Abel (2013) observes that in periods with adverse weather shocks, the prices of agricultural products rose more in cities without access to water borne transport. Heterogeneity in the land constraint is examined below.

[34] Basic land-intensive products account for 2/3's of spending in Allen's (2008) consumer price index.

[35] OLS estimates confirm the decline in correlation between prices and city size. The OLS estimate (standard errors in parentheses) declines from 0.21 (0.08) in 1600 to 0.19 (0.07) in 1700 and 0.17 (0.07) in 1800. In contrast with the positive historical correlation between city size and food prices, food prices are not systematically correlated with city size in leading economies today. For example, the US Bureau of Labor Statistics CPI for food consumed at home by urban wage earners (series CWURA000SAF11) is consistently highest for small cities (population less than 50 thousand) and lowest for mid-sized cities (population between 50 thousand and 1.5 million). Food prices in the largest US cities (population above 1.5 million) are between these extremes.

volumes in the international grain trade within Europe. Second, new trades routes, significant increases in productivity in mercantile and knowledge-based activities, and institutional change favorable to commerce all fostered growth in larger cities. Third, there were improvements in agricultural productivity.

The international grain trade transformed the economic relationships supporting city growth. The Baltic grain trade emerged as the first large scale international trade in basic wage goods in European history since the Roman era and was known to the Dutch as the *moedernegotie*: the "mother of all trades." The emergence of the Baltic trade in grain can be clearly dated to the mid-15th century (Davies 1981: 256). At this time, increasing urban demand from the Netherlands was met by increasing surpluses from the newly reunited Vistula River basin in Poland. In the Northern Netherlands, where city growth was unusually rapid, the bread and beer supply of city populations depended on grain imports. Grain came to be imported from Eastern Europe in quantities sufficient to feed 1 in 4 inhabitants of the Dutch Republic over the course of the 1500s (van Tielhof 2002: 1). For example, Dutch grain imports were sufficient to feed 260 thousand people in the mid-1500s and by the 1590s were sufficient to feed 436 thousand people – at a time when the total population of Dutch cities was 429 thousand (Bairoch et al. 1988 provide information on the top 29 cities; see also Appendix E). To be clear, considerable quantities of grain were re-expored from the Netherlands to Mediterranean: "grain was the commodity that gave Dutch merchants entrée to the Iberian and Mediterranean ports from the 1590s on" (de Vries and van der Woude 1997: 414-5). By the late 1500s, Iberian and Italian cities facing local resource constraints were dependent on imports of Baltic grain.

The development of large scale international trade in grain was associated with significant declines in freight shipping costs on maritime routes in the 16th century (van Zanden and Tielhof 2009; Menard 1991). In the 1510's, the cost of shipping rye from the Baltic to Amsterdam represented over 20 percent of its price when sold in Holland. By the 1580's and 1590's, ratio of shipping costs to sale prices fell to an average of 10 percent (see van Tielhof 2002: 198). Between the 1450s and the 1650s, indexed Dutch shipping costs fell by more than 50 percent (van Tielhof and van Zanden 2009). Over the same period, the cost of shipping wine from Bordeaux to London fell even further, declining by approximately 70 percent (Menard 1991). These developments were made possible by innovations in shipping technology, notably the introduction of the *fluyt* vessel, and the pacification of the Baltic (Menard 1991; van Tielhof and van Zanden 2009). Similar increases in productivity and declines in costs are observed in trans-Atlantic shipping. Menard (1991) finds productivity growth of more than 1% per year in shipping rice from Charleston, South Carolina to London 1700-1776 and similar rates of productivity growth in trans-Atlantic tobacco shipping over even longer periods.[36]

The period 1500-1700 was more broadly characterized by increased returns to mercantile and knowledge-based activities that concentrated in larger cities. The period between 1500 and 1750 was the pre-industrial "age of commercial Europe" (Minchinton 1974: 82). Atlantic ports benefitted from new trade routes connecting them to both the Atlantic and the East Indies (Acemoglu, Johnson, and Robinson 2005; Glamann 1974). The knowledge revolution associated with the diffusion of the printing press raised productivity in business, administration, and educational activities. Cities with printing presses grew as much as 60% faster than otherwise similar cities 1500-1600 (Dittmar 2011). Of all the bodies of knowledge in print, the ideas in the new business education literature were uniquely associated with city growth. Comprehensive firm-level microdata on all printing firms in all European cities supports the view that that the diffusion of knowledge drove growth. The microdata reveal that variations in city-level supply of business education content were induced by the chance deaths of heterogeneous printing firm managers, and that overall city growth responded to the variations in supply induced by manager deaths (Dittmar 2015). This evidence supports the view that ideas in the business education literature played a role in producing or activating upper tail entrepreneurial talent, but suggests that random local shocks contributed materially to this process and does not directly point to the emergence of particular distribution of talent imagined in talent-based models of Zipf's Law.

Increases in agricultural productivity also relaxed the land constraint, and are regarded by historians as a key factor in the growth of cities in Northwest Europe (Pounds 1990, Kriedte 1979). Agricultural productivity was positively associated with urbanization, and economies where city growth was concentrated experienced relatively high productivity growth in agriculture. Appendix E presents estimates of agricultural productivity at the national level and shows how advances in productivity were associated with increases in urbanization (Allen 2003). Narrative evidence strongly indicates that variation in agricultural productivity reflected the dynamics of commerce and trade: "the determining factor was the tremendous and disparate development of commercial capitalism" (Maddalena 1974: 276).

While the period 1500-1800 saw marked increases in agricultural productivity and in

---

[36] For details of the data, see Appendix E. Data for periods before 1450 are limited, in part because international transactions that emerged 1450-1600 were new trades. However, the available evidence suggests that the costs of shipping were relatively low in the early 1300s. Menard (1991) observes that for wine the "transport revolution" of the Renaissance may have only returned real freight rates to 14th century levels. This suggests that organizational and institutional factors also mattered – for example in Poland, the consolidation of an economic regime featuring coerced (serf) labor producing agricultural goods for export.

the trade in food, the public health environment in cities did not improve substantially until at least the 1700s. An extensive literature on the demography of early modern cities finds that urban death rates exceeded urban birth rates and rural death rates.[37] In general, mortality was increasing in city size. Other things equal, this limited the growth of large cities. However, the big revolutions in public health came after 1800 and mortality in large cities remained relatively high throughout the early modern period. The plague stopped striking the cities of Western Europe only in the 1700s – after city growth in Western Europe became size-independent.[38]

5.3 Mobility of and Returns to Upper Tail Human Capital

Historical evidence suggests that the returns to, and the distribution of, entrepreneurial talent and upper tail human capital were transformed during the early modern period by the dissemination of knowledge and changes in the conditions of trade. The evidence also indicates considerable labor mobility before Zipf's Law emerged, making it less plausible that Zipf's Law emerged because a migration channel "turned on" and enabled selection of upper tail talent across locations starting around 1500.

The early modern era was characterized by significant innovations in, and much wider diffusion of, the knowledge used by entrepreneurs (e.g. Braudel 1979a, b; Dittmar 2015). The new forms of knowledge related to the use of innovations in accounting, book-keeping, applied mathematics, contract design, and forms of business protocol and even comportment (Dittmar 2015). These innovations were disseminated in the emerging business education literature, informally, and via formal schooling (e.g. so-called *abbaco* schools that prepared youths for careers as merchants).

Historical evidence on changes in business practices suggests that the returns to upper tail talent were contingent on the diffusion of knowledge. If Zipf's Law reflects an approximate power law of entrepreneurial talent (Behrens, Duranton, and Nicoud 2014), the historical evidence suggests that the distribution of this entrepreneurial talent was likely itself the outcome of a larger process of the development of social knowledge. One possibility is that upper tail talent was produced. Another is that the diffusion of knowledge in effect activated upper tail talent that was previously latent.

---

[37] See Woods (2003), Mols (1955, 1956), de Vries (1984), Bairoch (1988).

[38] Arguably the big improvement in public health in early modern Europe was the virtual disappearance of the plague in the 1700s. While procedures for city-level plague quarantine date to before 1500, these procedures were difficult for municipal governments to fully enforce given their limited state capacity, were unable to eradicate reservoirs of plague within Europe, and do not appear to explain the timing of what Lindeman (1999, p. 64) observes is "the mystery surrounding the end of the plague in Europe."

Changes in the conditions of trade are also likely to have impacted the returns to talent. New trade routes to the Americas and Asia together with lower transport costs within Europe may have generated new opportunities for talented entrepreneurs, enabling super-star type returns to emerge. It is plausible that these changes worked to generate a thicker – approximately Pareto – upper tail in the returns to talent, however our quantitative evidence on the historical distribution of and returns to talent is exceedingly fragmentary. Changes in the conditions of trade may have also made it feasible for upper tail talent to emerge simply by relaxing the land constraint, as discussed above. For example, it is plausible that certain types of talents could only realize returns in large cities – with large markets or when able to profit from characteristics of information sharing and/or labor markets in large cities. A world in which city growth was constrained by the cost of transport and the limits of agricultural productivity, would then be one in which the upper tail of the effective talent distribution may have been thin rather than Pareto thick.

Before the emergence of Zipf's Law there was considerable labor mobility in Western Europe. The evidence on mobility is not definitive but casts doubt on the possibility that Zipf's Law emerged because previously immobile talent became mobile. Several pieces of evidence are notable. First, the expansion of urban Europe into Eastern Germany and Poland in the middle ages involved entrepreneurs establishing brand new cities, towns, and villages and devising settlement incentives that drew large numbers Western German and Dutch settlers over long distances (Aubin 1966). Second, municipalities maintained records of residents with formal municipal citizenship rights (e.g. *bürgerbücher* lists of burgers) and these indicate many cities had 50% of their citizen-residents with "foreign" origins (Hochstadt 1983, Mols 1954). Third, there is considerable evidence on the formal and informal circulation of skilled labor in Medieval and early modern Europe, which shows that the environment was characterized by considerable movement over distance for the purposes of receiving training and establishing businesses (Reith 2004).

6 Institutions and Cities in Eastern Europe

Comparisons between Western and Eastern Europe highlight the role of institutions in shaping early modern city growth and the emergence of Zipf's Law. During the early modern era, economies in Eastern Europe established institutions restricting the mobility of labor and economic activity in cities.[39] In Eastern Europe new legal institutions estab-

---

[39] See Bogucka (1982), Pachs (1994a), Berend (1986), Süchs (1988), Kahan (1973), Anderson (1974a, 1974b), Melton (1988), Blum (1957, 1978), Carsten (1954), Brenner (1976), and Topolski (1982).

lished “major extensions and intensifications of serfdom” and reflected, “concerted action to restrict the rights and mobility of the peasantry” (Bideleux and Jeffries 2007: 161). These laws prevented tenant farmers from leaving tenancies without the formal consent of landlords, established restrictions on seasonal migration, and set-up mechanisms for enforcement backed by corporal punishment. Historical research suggests that these institutions impacted the development of cities in Eastern Europe. These restrictions were lifted starting in the early 1800s, with the abolition of serfdom in Eastern Europe (Blum 1957, 1978). Did city population distributions and the underlying pattern of city growth evolve differently where institutions restricting labor mobility were established?

Figure 9 graphs the distribution of city sizes and shows that deviations from Zipf’s Law were relatively persistent in Eastern Europe and of two kinds. First, we observe a pronounced curvature in the the log rank-log size plots through 1700, consistent with non-random growth in which small cities grow relatively quickly (see Rossi-Hansberg and Wright 2007). Second, in 1800 we observe that the four largest cities of Eastern Europe were substantially *bigger* than we would predict based on the rest of the city size distribution. These cities are the capitals: Berlin, Vienna, St. Petersburg, and Moscow. The relative size of these “urban giants” reflects the fact that they enjoyed extremely fast growth 1700 to 1800 and is consistent with contemporary evidence on cities which enjoy politically generated rents and as a result grow to be larger than expected given the size of other cities in the economy (Ades and Glaeser 1995).

Figure 10 plots the distribution of city growth rates by city population quintile and shows that growth in Eastern Europe remained size dependent longer than in the West. In Eastern Europe, smaller cities grew relatively quickly through the 1600s. The growth deficit in the upper tail only ended in the 1700s, when political capitals in Eastern Europe experienced relatively high growth.

The Eastern Europe comparison provides an institutional contrast that operates principally in the cross-section dimension. This can be seen as a complement to the comparisons between city populations in Western Europe in the early modern era and under the Roman empire – which highlight institutional differences in the longer time series, as noted above. Taken together with historical evidence on food security and the international grain trade, these comparisons underline that the nature of market relationships and the technological determinants of city populations cannot be considered separately from the institutions that shaped the pattern of urbanization.

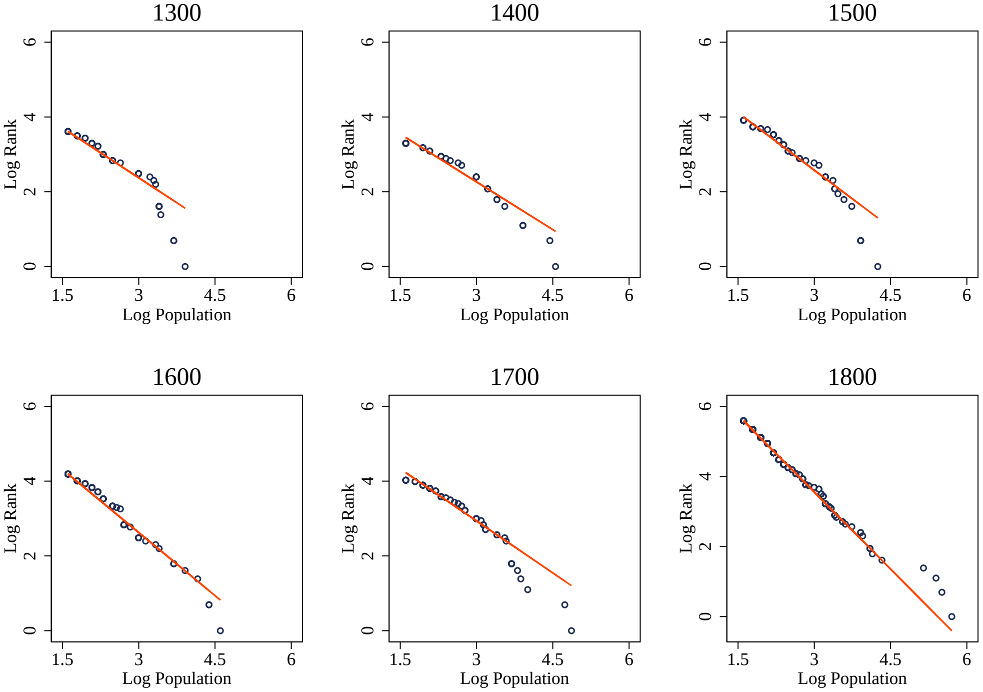
7 Conclusion

Zipf's Law is one of the most robust empirical regularities in economics. The apparent ubiquity of Zipf's Law raises the possibility that fundamental determinants of the location of economic activity are trans-historical and may not respond to changes in institutions and technology commonly viewed as marking significant shifts in economic history.

This research documents that Zipf's Law emerged over time in European history. In the middle ages, relatively large cities struggled to grow and remained too small to satisfy Zipf's Law. Zipf's Law emerged between 1500 and 1800, after city growth rates became random, in the sense of being independent of initial city population. This transformation in Europe's urban system occurred during the period of transition to modern economic growth and before the industrial revolution.

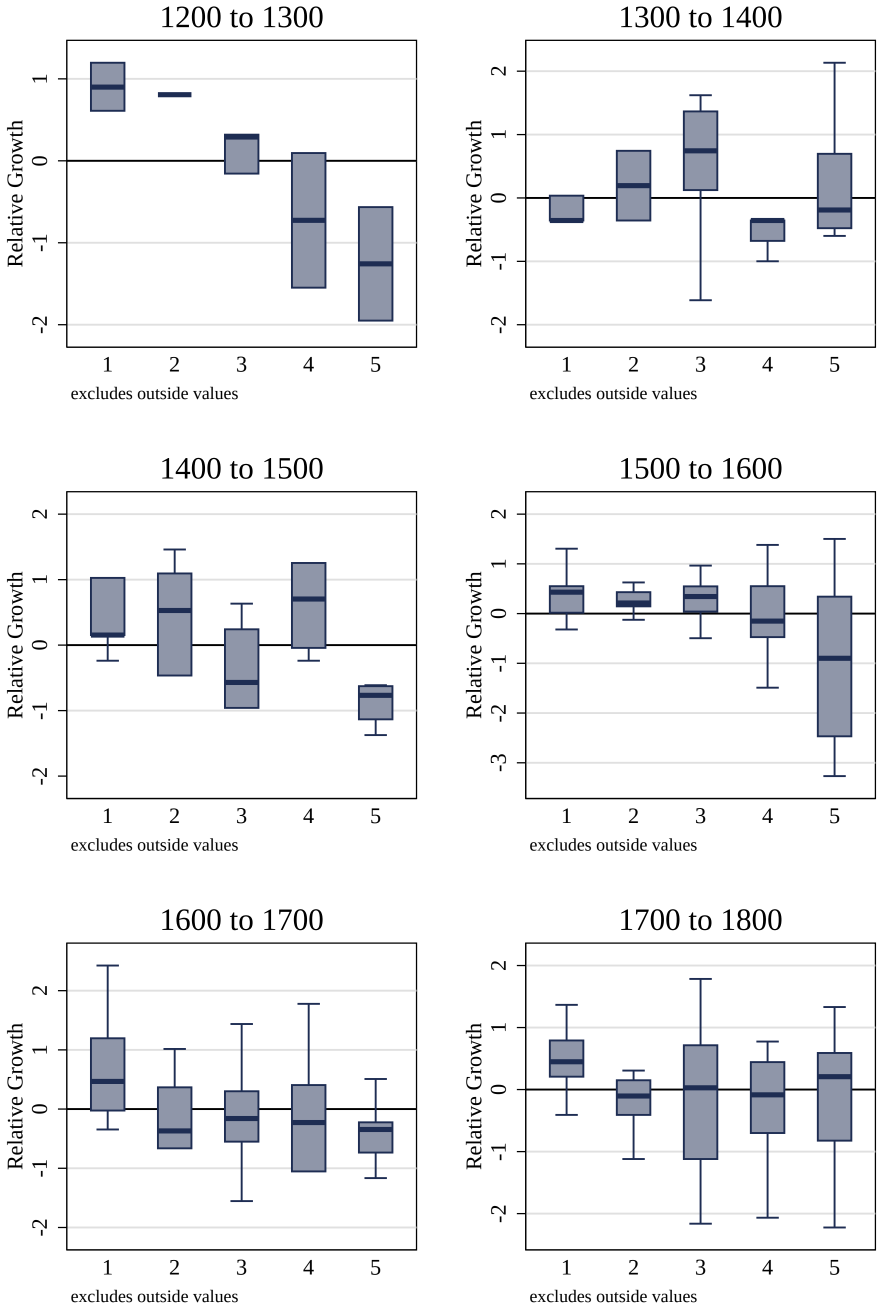
The historical emergence of Zipf's Law has implications for economic theory. The fact that Zipf's Law emerged over time – while the principal features of the landscape were invariant – suggests that narrowly geographic explanations will be insufficient. Propitious locations are non-homogeneous and distributed unevenly, but the historical emergence of

Figure 9: The Emergence of Zipf's Law in Eastern Europe



This figure plots (1) raw data on city populations (S_i) and their corresponding size rankings (R_i), and (2) fitted values estimated using robust non-parametric Theil regression and the model: $\ln(R_i) = \alpha - \beta \ln(S_i) + \epsilon_i$. Populations in thousands are from Bairoch, Batou, and Chèvre (1988).

Figure 10: The Distribution of City Growth Rates by Size Quantile in Eastern Europe



This graph presents the distribution of city growth rates in Eastern European cities by initial population quintile between 1000CE and 1800CE. In each period, the smallest cities are in quintile 1 and the largest are in quintile 5. The boxes describe the interquartile range. The line within each box is that quintile's median. The "whiskers" mark the adjacent values. Growth rates are normalized in standard deviation units. If the population growth rate of city i is g_{it} in period t, and the mean and standard deviation across cities are $\bar{g}_t$ and σ_t, then normalized growth is $\hat{g}_{it} = (g_{it} - \bar{g}_t)/\sigma_t$.

Zipf's Law suggests that locational advantages may emerge with economic development, and hence be endogenous along important dimensions. The fact that Zipf's Law emerged in an era when the industrial specialization of urban activity was relatively limited also suggests that explanations emphasizing cities specialized in the production of particular goods and reaching optimal size for their activity may not capture the root process.

Historically, Zipf's Law emerged in an era characterized by the diffusion of knowledge-based innovations and changes in international trade – including Europe's first large scale international trade in foodstuffs since the Roman era. The fact that large cities had been possible in the Roman era highlights the fact that city size distributions reflect institutional environments. In the early modern era, laws restricting labor mobility in Eastern Europe were associated with persistent deviations from Zipf's Law. In Western Europe, where restrictions on mobility were relatively relaxed and where the benefits of new patterns of trade were most pronounced, Zipf's Law emerged.

References

[1] Abel, Wilhelm (2013), *Agricultural Fluctuations in Europe: From the Thirteenth to Twentieth Centuries*, Routledge; New York.

[2] Acemoglu, Daron, Simon Johnson, and James Robinson (2005), "The Rise of Europe," *American Economic Review*, Vol. 95, No. 3, pp. 546-579.

[3] Acemoglu, Daron, Simon Johnson, and James Robinson (2002), "The Rise of Europe," NBER Working Paper 9378.

[4] Ades, Alberto and Glaeser, Edward (1995), "Trade and circuses: explaining urban giants," *Quarterly Journal of Economics*,Vol. 110, pp. 195-227.

[5] Allen, Robert (2001), "The Great Divergence in European Wages and Prices from the Middle Ages to the First World War," *Explorations in Economic History*, Vol. 38, pp. 411-447.

[6] Allen, Robert (2003), "Progress and poverty in early modern Europe," *Economic History Review*, Vol. LVI, No. 3, pp. 403-443.

[7] Allen, Robert (2008), "Consumer price indices, nominal / real wages and welfare ratios of building craftsmen and labourers, 1260-1913," available on-line: http://www.iisg.nl/hpw/data.php.

[8] Anderson, Perry (1974a), *Passages From Antiquity to Feudalism*, London; Verso.

[9] Anderson, Perry (1974b), *Lineages of the Absolutist State*, London; Verso.

[10] Aubin, Hermann (1966), "The lands east of the Elbe and German colonization eastwards," in M. Postan ed., *The Cambridge Economic History of Europe*, Vol. I, Cambridge; Cambridge University Press.

[11] Bairoch, Paul (1988), *Cities and Economic Development*, Chicago; University of Chicago.

[12] Bairoch, Paul, Jean Batou, and Pierre Chèvre (1988), *La Population Des Villes Européenes*, Geneva; Droz.

[13] Ballaux, B. and B. Blondé (2004), "Transport Prices in the Long Sixteenth Century," on-line at: www.lowcountries.nl/papers/2004-4_ballaux.pdf

[14] Bath, Slicher van (1977), "Agriculture in the Vital Revolution," in E. Rich and C. Wilson eds., *The Cambridge Economic History of Europe*, Vol. 5, Cambridge; Cambridge University.

[15] Behrens, Kristian, Gilles Duranton, and Frédéric Robert-Nicoud (2014), "Productive cities: Sorting, selection and agglomeration," *Journal of Political Economy*, Vol. 122, No. 3, pp. 507-553.

[16] Behrens, Kristian and Frédéric Robert-Nicoud (2015), "Agglomeration Theory with Heterogeneous Agents," in Gilles Duranton, J. Vernon Henderson, and William Strange eds., *Handbook of Regional and Urban Economics*, Vol. 5A, North-Holland.

[17] Berend, Ivan (1986), "The historical evolution of Eastern Europe as a region," *International Organization*, Vol. 40, No. 2, pp. 329-346.

[18] Berman, Harold (1983), *Law and Revolution, the Formation of the Western Legal Tradition*, Cambridge; Harvard University Press.

[19] Berman, Harold and Colin Kaufman (1978), "Law of International Commercial Transactions (Lex Mercatoria)," *Harvard International Law Journal*, Vol. 19, No. 1, ppp. 221-277.

[20] Bideleux, Robert and Ian Jeffries (2007), *A History of Eastern Europe*, London; Routledge.

[21] Black, Duncan and J.V. Henderson (1999), "A Theory of Urban Growth," *Journal of Political Economy*, Vol. 107, No. 2.

[22] Bleakley, Hoyt and Jeffrey Lin (2012), "Portage and Path Dependence," *Quarterly Journal of Economics*, Vol. 127, No. 2, pp. 587-644.

[23] Blockmans, Wim (1994), "State Formation in Preindustrial Europe," in Charles Tilly and Wim Blockmans eds., *Cities and the Rise of States in Europe, AD 1000 to 1800*, Boulder, CO; Westview.

[24] Blum, Jerome (1957), "The Rise of Serfdom in Eastern Europe," *American Historical Review*, Vol. 62, No. 4, pp. 807-836.

[25] Blum, Jerome (1978), *The End of the Old Order in Rural Europe*, Princeton; Princeton University.

[26] Bogucka, Maria (1982), "Polish towns between the sixteenth and eighteenth centuries," in J. Fedorowicz et al. eds., *A Republic of Nobles: Studies in Polish History to 1864*, Cambridge; Cambridge University.

[27] Bosker, Maarten, Eltjo Buringh, and Jan Luiten van Zanden (2008), "From Baghdad to London: The Dynamics of Urban Growth in Europe and the Arab World, 800-1800," CEPR Discussion Paper DP6833.

[28] Botero, Giovanni (1602), *A Treatise Concerning the Causes of the Magnificency and Greatness of Cities*: http://socserv.mcmaster.ca/econ/ugcm/3ll3/botero/cities.

[29] Brakman, Steven, Harry Garretsen, and Marc Schramm (2004), "The strategic bombing of German cities during WWII and its impact on city growth," *Journal of Economic Geography*, Vol. 4, pp. 201-218.

[30] Braudel, Fernand (1966), *La Méditerranée et le Monde Méditerranéen À L'Époque De Phillippe II*, Paris; Armand Colin.

[31] Braudel, Fernand (1979a, 1979b, 1979c), *Civilisation, Économie et Capitalisme*, Vols. I, II, and III, Paris; Armand Colin.

[32] Brenner, Robert (1976), "Agrarian Class Structure and Economic Development in Pre-Industrial Europe," *Past and Present*, Vol. 70, pp. 30-75.

[33] Brinkhoff, Thomas (2008), "City Population": http://www.citypopulation.de/. Data accessed March 2008.

[34] Cantoni, Davide, Jeremiah Dittmar, and Noam Yuchtman (2015), "Reformation and Reallocation: Human Capital, Employment, and Economic Growth in the German Lands," working paper.

[35] Carsten, Francis (1954), *The Origins of Prussia*, Oxford; Clarendon.

[36] Chittolini, Giorgio (1994), "City-States and Regional States in North-Central Italy," in Charles Tilly and Willem Blockmans eds., *Cities and the Rise of States in Europe*, Westview; Boulder.

[37] Cipolla, Carlo (1982), *Before the Industrial Revolution: European Society and Economy 1000-1700*, 2nd Edition, London; Methuen.

[38] Cook, Chris and John Stevenson (1978), *Longman Atlas of Modern British History*, London; Longman.

[39] Cordoba, Juan (2008), "On the Distribution of City Sizes," *Journal of Urban Economics*, Vol. 63, Issue 1, pp. 177-197.

[40] Darby, H.C. ed. (1970), *The New Cambridge Modern History*, Vol. XIV, Cambridge; Cambridge University Press.

[41] Davies, Norman (1981), *God's Playground: A History of Poland*, New York; Columbia University.

[42] Davis, Donald and David Weinstein (2002), "Bones, Bombs, and Breakpoints: the Geography of Economic Activity," *American Economic Review*, Vol. 92, No. 5, pp. 1269-1289.

[43] DeLong, J. Bradford and Andrei Shleifer (1993), "Princes and Merchants: City Growth Before the Industrial Revolution," *Journal of Law and Economics*, Vol. 36, October, pp. 671-702.

[44] Desmet, Klaus and and J. Vernon Henderson (2015), "The Geography of Development within Countries," in Gilles Duranton, J. Vernon Henderson, and William Strange eds., *Handbook of Regional and Urban Economics*, Vol. 5B, North-Holland.

[45] Desmet, Klaus and Jordan Rappaport (2015), "The settlement of the United States, 1800-2000: The long transition towards Gibrat's Law," *Journal of Urban Economics*, forthcoming (in press).

[46] Dincecco, Mark and Massimiliano Ornato (2014), "Military Conflict and the Economic Rise of Urban Europe," SSRN working paper: `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2341713`

[47] Dittmar, Jeremiah (2011), "Information Technology and Economic Change: The Impact of the Printing Press," *The Quarterly Journal of Economics*, Vol. 126, No. 3, pp. 1133-1172.

[48] Dittmar, Jeremiah (2015), "New Media, Competition, and Growth: European Cities After Gutenberg," www.jeremiahdittmar.com/research

[49] Drennan, R. and C. Peterson (2004), "Comparing achaeological settlement systems with rank-size graphs: a measure of shape and statistical confidence," *Journal of Archaeological Science*, No. 31, pp. 533-549.

[50] Eeckhout, Jan (2004), "Gibrat's Law for (All) Cities," *American Economic Review*, Vol. 94, No. 5.

[51] Eurostat (2009), "Total population" (tps00001), on-line at: http://epp.eurostat.ec.europa.eu/tgm/table.do?tab=table&language=en&pcode=tps00001&tableSelection=1&footnotes=yes&labeling=labels&plugin=1

[52] Friedrichs, Christopher (1995), *The Early Modern City, 1450-1750*, London; Longman.

[53] Gabaix, Xavier (1999a), "Zipf's Law for Cities: An Explanation," *Quarterly Journal of Economics*, Vol. 114, No. 3, pp. 739-767.

[54] Gabaix, Xavier (1999b), "Zipf's Law and the Growth of Cities," *American Economic Review*, Papers and Proceedings, Vol. 89, No. 2, pp. 129-132.

[55] Gabaix, Xavier (2009), "Power Laws in Economics and Finance," *Annual Reviews of Economics*, Vol. 1, pp. 255-294.

[56] Gabaix, Xavier (2016), "Power Laws in Economics: An Introduction," *Journal of Economic Perspectives*, Vol. 30, No. 1, pp. 185206.

[57] Gabaix, Xavier and Yannis Ioannides (2004), "The Evolution of City Size Distributions," in J. Henderson and J. Thisse eds., *The Handbook of Regional and Urban Economics*, Vol. 4.

[58] Gabaix, Xavier and Rustem Ibragimov (2011), "Rank-1/2: A Simple Way to Improve the OLS Estimation of Tail Exponents," *Journal of Business and Economics Statistics*, Vol. 29, No. 1, pp. 24-39.

[59] Gibson, Campbell and Jay Jung (2005), "Historical Census Statistics On Population Totals By Race, 1790 to 1990, and By Hispanic Origin, 1970 to 1990, For Large Cities And Other Urban Places In The United States," U.S. Census Bureau, Population Division, Working Paper No. 76.

[60] Glamann, Kristof (1974), "European Trade 1500-1750," in Carlo Cipolla ed., *The Fontana Economic History of Europe*, Fontana; Glasgow.

[61] Graham, Brian (1979), "The Evolution of Urbanization in Medieval Ireland," *Journal of Historical Geography*, Vol. 5, No. 2, pp. 111-125.

[62] Greif, Avner (2006), *Institutions and the Path to the Modern Economy: Lessons from Medieval Trade*, Cambridge; Cambridge University Press.

[63] He, Xuming, Jana Jureckova, and Roger Koenker (1990), "Tail Behavior of Regression Estimators and their Breakdown Points," *Econometrica*, Vol. 58, No. 5, pp. 1195-1214.

[64] Hellie, R. (1971), *Enserfment and Military Change in Muscovy*, Chicago; University of Chicago.

[65] Henderson, J.V. (1974), "The Sizes and Types of Cities," *American Economic Review*, Vol. 61, pp. 640-656.

[66] Hochstadt, Steve (1983), "Migration in Preindustrial Germany," *Central European History*, Vol. 16, No. 3.

[67] Hohenberg, Paul M. and Lynn H. Lees (1985), *The Making of Urban Europe, 1000-1950*, Cambridge; Harvard University.

[68] Hollander, Myles and Douglas Wolfe (1999), *Nonparametric Statistical Methods*, 2nd edition, New York; Wiley and Sons.

[69] India Office (1905), *Statistical Abstract Relating to British India from 1894-95 to 1903-04*, London; His Majestie's Stationery Office.

[70] Ioannides, Yannis and Harry Overman (2003), "Zipf's Law for cities: an empirical examination," *Regional Science and Urban Economics*, Vol. 33, pp. 127-137.

[71] Ioannides, Yannis et al., "The Effect of Information and Communication Technologies on Urban Structure," *Economic Policy*, April, pp. 201-242.

[72] Ioannides, Yannis and Junfu Zhang (2015), "Walled Cities in Late Imperial China," Tufts University Working Paper.

[73] Johnson, Gregory (1980), "Rank-Size Convexity and System Integration: A View from Archaeology," *Economic Geography*, Vol. 56, No. 3, pp. 234-247.

[74] Jones, Eric (1981), *The European Miracle: Environments, economics, and geopolitics in the history of Europe and Asia*, Cambridge; Cambridge University Press.

[75] Kahan, Arcadius (1973), "Notes on Serfdom in Western and Eastern Europe," *Journal of Economic History*, Vol. 33, No. 1, pp. 86-99

[76] Kamiński, Andrzej (1975), "Neo-Serfdom in Poland-Lithuania," *Slavic Review*, Vol. 34, No. 2, pp. 253-268.

[77] Koenker, Roger (2005), *Quantile Regression*, Cambridge; Cambridge University.

[78] Kriedte, Peter (1979), *Peasants, Landlords and Merchant Capitalists: Europe and the World Economy 1500-1800*, Cambridge; Cambridge University.

[79] Krugman, Paul (1996a), *The Self-Organizing Economy*, Cambridge, MA; Blackwell.

[80] Krugman, Paul (1996b), "Confronting the Urban Mystery," *Journal of the Japanese and International Economies*, Vol 10, pp. 399-418.

[81] Lee, Sanghoon, and Qiang Li (2013), "Uneven Landscapes and City Size Distributions," *Journal of Urban Economics*, Vol. 78, p. 19-29.

[82] Lindeman, Mary (1999), *Medicine and Society in Early Modern Europe*, Cambridge; Cambridge University.

[83] Livet, Georges (2003), *Histoire des routes et des transports en Europe: Des chemins de Saint-Jacques*, Strasbourg; Presses Universitaires de Strasbourg.

[84] Livi-Bacci, Massimo (1999), *The Population of Europe*; Oxford, Blackwell.

[85] Maddalena, Aldo de (1977), "Rural Europe 1500-1750," in Carlo Cipolla ed., *The Fontana Economic History of Europe: The 16th and 17th Centuries*, London; Fontana.

[86] Magosci, Paul (1993), *Historical Atlas of East Central Europe*, Seattle; University of Washington.

[87] Makkai, Lázló (1975), "Neo-Serfdom: Its Origin in East Central Europe," *Slavic Review*, Vol. 34, No. 2, pp. 225-238.

[88] Meiggs, Russell (1973), *Roman Ostia*, Oxford; Clarendon.

[89] Melton, Edgar (1988), "*Gutherrschaft* in East Elbian Germany and Livonia, 1500-1800: A Critique of the Model," *Central European History*

[90] Menard, R. R. (1991), 'Transport costs and long-range trade, 1300-1800: was there a European 'transport revolution' in the early modern era?", in Tracy, J. D., ed., *The political economy of merchant empires*, Cambridge; Cambridge University Press.

[91] Michaels, Guy, Ferdinand Rauch, and Stephen Redding (2012), "Urbanization and Structural Transformation," *Quarterly Journal of Economics*, Vol. 127, No. 2, pp. 535-585.

[92] Minchinton, Walter (1974), "Patterns and Structure of Demand 1500-1750," in Carlo Cipolla ed., *The Fontana Economic History of Europe*, Fontana; Glasgow.

[93] Miskimin, Harry (1975), *The Economy of Early Renaissance Europe, 1300-1460*, Cambridge; Cambridge University Press.

[94] Mols, Roger (1954, 1955, 1956), *Introduction à la Démographie Historique des Villes D'Europe du XIV^{e} au XVIII^{e} siècle*, Tomes 1-3, Louvain; Éditions J. Duculot/Publications Universitaires de Louvain.

[95] Morris, Ian (2013), *The Measure of Civilisation: How Social Development Decides the Fate of Nations*, London; Profile.

[96] Nicholas, David (2003), *Urban Europe, 1100-1700*, London; Palgrave MacMillan.

[97] Nunn, Nathan and Nancy Qian (2011), "The Potato's Contribution to Population and Urbanization: Evidence from a Historical Experiment," *Quarterly Journal of Economics*, Vol. 126, No. 2, pp. 593-650.

[98] Pach, Zigismund (1994a), *Hungary and the European Economy in Early Modern Times*, Aldershot, UK; Varorium.

[99] Pach, Zigismund (1994b), "The East-Central European aspect of the overseas discoveries and colonization," in H. Pohl ed., *The European Discovery of the World and its Economic Effets on Pre-Industrial Society 1500-1800*, Stuttgart; Franz Steiner.

[100] Pirenne, Henri (1927) [1969 reprint], *Medieval Cities: Their Origins and the Revival of Trade*, Princeton; Princeton University.

[101] Postan, Michael M. (1979), *The Medieval Economy and Society*, London; Penguin.

[102] Pounds, Norman (1990), *An Historical Geography of Europe*, Cambridge; Cambridge University.

[103] Pounds, Norman (1979), *An Historical Geography of Europe, 1500-1840*, Cambridge; Cambridge University.

[104] Reith, Reinhold (2008), "Circulation of Skilled Labour in Late Medieval and Early Modern Central Europe," in S.R. Epstein and Maarten Prak eds., *Guilds, Innovation and the European Economy, 1400-1800*, Cambridge; Cambridge University Press.

[105] Robisheaux, Thomas (1998), "The Peasantries of Western Germany, 1300-1750," in Tom Scott ed., *The Peasantries of Europe from the Fourteenth to the Eighteenth Centuries*, London; Longman.

[106] Rossi-Hansberg, Esteban and Mark Wright (2007), "Urban Structure and Growth," *Review of Economic Studies*, Vol. 74, pp. 597-624.

[107] Rousseeuw, Peter and Annick Leroy (1987), *Robust Regression and Outlier Detection*, New York; John Wiley.

[108] Rowley, Trevor, *The English Landscape in the Twentieth Century*; London, AC Black.

[109] Rozenfeld, Hernán, Diego Rybski, Xavier Gabaix, and Hernán Makse (2011), "The Area and Population of Cities: New Insights from a Different Perspective on Cities," *American Economic Review*, Vol. 101, No. 5, pp. 2205-2225.

[110] Russell, Josiah (1972), *Medieval Regions and Their Cities*, Bloomington; Indiana University.

[111] Samsonowicz, Henryk and Antoni Mączak (1985), "Feudalism and capitalism: a balance of changes in East-Central Europe," in Peter Burke ed., *East-Central Europe in transition*, Cambridge; Cambridge University.

[112] Savage, S. (1997), "Assessing Departures from Log-Normality in the Rank-Size Rule," *Journal of Archaeological Science*, No. 24, pp. 233-244.

[113] Scott, Tom (2005), *Town, Country, and Regions in Reformation Germany*, Leiden; Brill.

[114] Simon, Henry (1955), "On a Class of Skew Distribution Functions," *Biometrika*, Vol. 42, No. 3/4, pp. 425-440.

[115] Smith, Carol (1990), "Types of City-Size Distributions," in A. van der Woude et al. eds., *Urbanization in History: A Process of Dynamic Interactions*, Oxford; Clarendon Press.

[116] Smith, Carol (1982), "Modern and premodern urban primacy," *Comparative Urban Research*, Vol. 11, pp. 79-96.

[117] Soo, Kwok Tong (2005), "Zipf's Law for cities: a cross-country investigation," *Regional Science and Urban Economics*, Vol. 35, pp. 239-263.

[118] Stabel, Peter (2008), "Composition et Recomposition des Réseaux Urbains des Pays-Bas," in Élizabeth Crouzet-Pavan and Élodie Lecuppre-Desjardin eds., *Villes de Flandre et Italie XIIIe-XVIe siècles*, Turnhout; Brepols.

[119] Statistics UK (2001), *Census 2001*, on-line at:
http://www.statistics.gov.uk/census2001/pop2001/northern_ireland.asp.

[120] Stillwell, Richard (1976), *Princeton Encyclopedia of Classical Sites*, Princeton; Princeton University.

[121] Stoianovich, Traian (1994), "Cities, Capital Accumulation, and the Ottoman Balkan Command Economy, 1500-1800," in Charles Tilly and Wim Blockmans eds., *Cities and the Rise of States in Europe, AD 1000 to 1800*, Boulder, CO; Westview.

[122] Süchs, Jenö (1988), "Three Historical Regions of Europe," in John Keane ed., *Civil Society and the State*, London; Verso.

[123] Sugar, Peter (1977), *A History of Southeastern Europe Under Ottoman Rule, 1354–1804*, Seattle; University of Washington.

[124] Szelényi, Balázs. (2006), *The Failure of the Central European Bourgeoisie: New Perspectives on Hungarian History*, New York; Palgrave.

[125] Theil, Henri (1950), "A rank-invariant method of linear and polynomial regression analysis," *Proceedings of the Royal Netherlands Academy of Sciences*, vol. 53, pp. 386-392.

[126] Topolski, Jerzy (1982), "Sixteenth-century Poland and the turning point in European economic development," in J. Fedorowicz et al. eds., *A Republic of Nobles: Studies in Polish History to 1864*, Cambridge; Cambridge University.

[127] Tielhof, Milja van (2002), *The "Mother of All Trades": The Baltic Grain Trade in Amsterdam from the Late Sixteenth to the Early Nineteenth Century*, Leiden; Brill.

[128] de Vries, Jan (1976), *The Economy of Europe in an Age of Crisis, 1600-1750*, Cambridge; Cambridge University Press.

[129] de Vries, Jan (1984), *European Urbanization 1500-1800*, Cambridge; Harvard University Press.

[130] de Vries, Jan (1990), "Problems in the Measurement, Description, and Analysis of Historical Urbanization," in A. van der Woude et al. eds., *Urbanization in History: A Process of Dynamic Interactions*, Oxford; Clarendon Press.

[131] de Vries, Jan and A.M. van der Woude (1997), *The First Modern Economy*, Cambridge; Cambridge University.

[132] Weber, Max (1958), *The City*, Glencoe, IL; Free Press.

[133] Woods, Robert (2003), "Urban-Rural Mortality Differentials: An Unresolved Debate," *Population and Development Review*, Vol. 29, No. 1, pp. 29-46.

[134] Zanden, Jan Luiten van and Milja Tielhof (2009), "Roots of growth and productivity change in Dutch shipping industry, 1500-1800," *Explorations in Economic History*, Vol. 46, pp. 389-403.

Appendices: Not For Publication

A Data

Data on city populations: Historic city populations are from Bairoch et al. (1988) and de Vries (1984). Contemporary city populations are from Brinkhoff (2008). City locations are from Bairoch et al. (1988) and http://www.batchgeocode.com/.

Data on Geography: Cities on navigable water comprised of the following

- Cities located on historic sea or ocean ports. Data on the historical location of ports are from Acemoglu, Johnson, and Robinson (2005), supplemented by data in Magosci (1993) and Stillwell (1976), and the sources cited in Dittmar (2011).
- Cities located on historically navigable in-land waterways. Data on navigable rivers are drawn from Magosci (1993), Pounds (1979, 1990), Livet (2003), Cook and Stevenson (1978), Graham (1979), Stillwell (1976), and de Vries and van der Woude (1997). The coding captures the principal historically navigable waterways, and does not class as "navigable" waterways that required substantial improvements and became navigable only over the early modern era. Some cities on navigable rivers were also sea or ocean ports.

Data on Laws Restricting Labor Mobility in Eastern Europe: Data on laws restricting labor mobility are from Data on the dates and and nature of the laws restricting labor mobility and limiting the rights of urban groups under the second serfdom is from Makkai (1975), Topolski (1982), Blum (1957), Carsten (1954), Szelényi (2006), Davies (1981), Pachs (1994), Hellie (1971), Kahan (1973), Kamiński (1975), Bogucka (1982), Melton (1988), Maddalena (1977), and Bideleux and Jeffries (2007).

B Measurement Error

I test for measurement error in the historical population data several ways. I first compare the Bairoch data to the most comprehensive independent source for city population data, the database in de Vries (1984). The Bairoch data covers all European cities that reached 5,000 inhabitants by or before 1800, has rich data from 1300 to 1850, and contains observations on 2,204 cities. The data in de Vries (1984) covers cities that reached a population of 10,000 between 1500 and 1800. It contains observations on 379 cities.

B.1 Ruling Out Measurement Error as the Explanation

This section discusses why observed deviations from Zipf's Law cannot be plausibly accounted for by non-classical measurement error or missing data.[40] I also present evidence showing that deviations from Zipf's Law are not a figment of aggregation, and are also observed in historical city size distributions at the national and sub-national levels.

I document that measurement error cannot account for the deviations from Zipf's Law by showing that the population short-falls in the upper tail are so big that they cannot be due to undercounting.[41] I show that if missing or mismeasured data for small cities were to account for observed deviations from Zipf's Law this would imply implausibly high urbanization rates 1300-1500. The observed deviations from Zipf's Law are consistent with a rich evidence on the historical demography of European cities as discussed in Section 5 below.

To gauge the possibility that the data undercount populations in the largest cities, I estimate hypothetical Zipf's Laws, calculate deviations from these benchmarks, and ask the question: How much larger (smaller) would outlier cities need to be to generate a pure log-linear relation? There are several reasons to use a robust regression estimator in this exercise. When data are generated by a stochastic power law, OLS estimators exhibit pronounced small sample bias (see Appendix C). Moreover, there appear to be outliers and the performance of OLS estimators is poor when there are heavy-tailed error distributions or when leverage points are present. Examination of the residuals from a robust regression can identify outliers (Koenker 2005).

Table 7 presents results for the largest cities in 1500 and shows that the magnitudes of the big city population shortfalls are so large that non-classical measurement error is not a plausible explanation for the observed deviations from Zipf's Law. Table 7 uses the Theil estimator to construct a measure of deviations from Zipf's Law. It compares observed population to the Theil regression predictions displayed in Figure 4. It documents that Paris and Naples would have needed to be three times larger than they were in 1500 – and, counterfactually, as big as they were in 1800 – to conform to the estimated hypothetical Zipf's Law. Table 7 also shows that if anything the short-falls are even larger in the data

---

[40] As noted above, classical error will not account for the observed deviations from Zipf's Law.

[41] As shown in Appendix B, a comparison of Bairoch data to the data in de Vries (1984) reveals no evidence that big city populations are systematically undercounted in the Bairoch data.

collected by de Vries (1984), which are examined in greater detail in the Appendix. The Appendix also presents similar results for other years and shows that between 1500 and 1700 the biggest cities were far smaller than they needed to be to satisfy a rank-size rule, consistent with the narrative evidence on historic city sizes presented in the next section.

Missing or mismeasured data on the populations of small cities also cannot account for the observed deviations from Zipf's Law. Cities with 10,000 inhabitants were substantial agglomerations unlikely to go missing in the data, and the observed deviations from Zipf's Law are robust to using a population cut-off of 10,000 instead of 5,000. Moreover, the argument that missing or mismeasured small cities account for deviations from Zipf's necessarily implies implausible, counterfactually high levels of urbanization 1300-1500. For instance, if we believe that the populations of the top 10 cities are correctly measured and that a power law holds, we can infer the implied populations of all subsequent cities in the urban hierarchy. This exercise implies European unobserved urbanization rates in 1300 that were implausibly far higher than observed urbanization rates in 1700.[42]

Table 7: The Populations of Europe's Largest Cities in 1500

(1) City	(2) Predicted Population	(3) Observed Bairoch	(4) Observed de Vries
Paris	690	225	100
Naples	378	125	150
Venice	266	100	100
Milan	266	100	100
Grenada	171	70	70
Lisbon	146	65	30
Tours	128	60	—
Genoa	114	58	60
Palermo	103	55	55
Gent	103	55	40

This table presents city populations in 1500 for the ten largest cities as recorded in Bairoch et al. (1988). Population figures are given in thousands. Column 2 presents the the predicted population values from Theil regression presented in Figure 4. Column 3 presents population observed in Bairoch et al (1988). Column 4 presends population recorded in de Vries (1984).

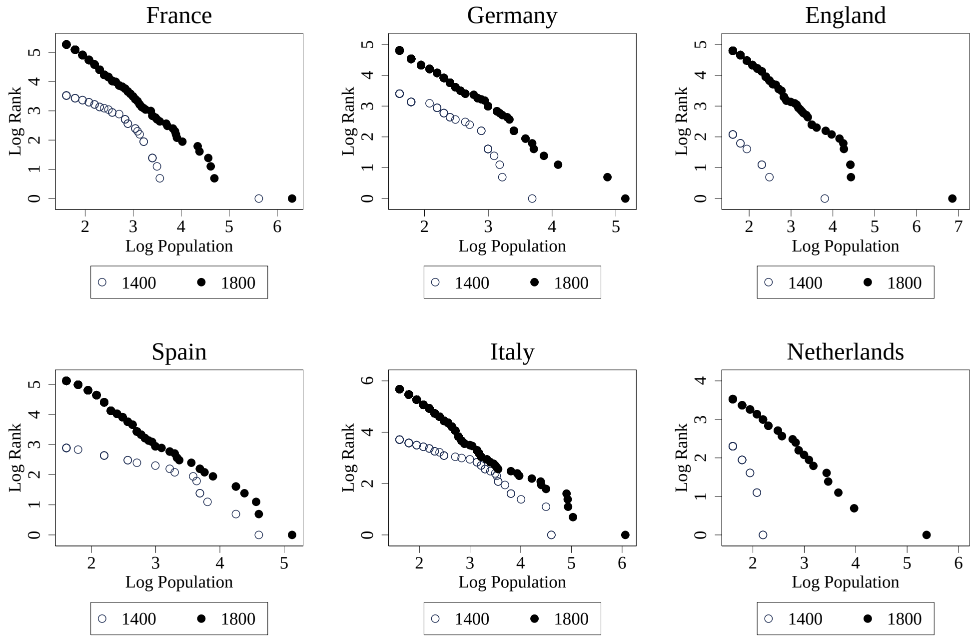
Deviations from Zipf's Law in the upper tail are also not a figment of the aggregation

---

[42] Imagine we examine only the set of cities with at least 5,000 inhabitants observed in 1300. To satisfy Zipf's Law all but the largest of these cities would need to be larger than observed in order to satisfy Zipf's Law. If the small cities were sufficiently large to satisfy the hypothetical power law implied by the populations of the largest 10 cities, that alone would deliver an urbanization rate of 14% in 1300 – equal to the rate observed in 1700. Moreover, other smaller cities with observed populations below 5,000 would also presumably need to be larger than observed to satisfy Zipf's Law, delivering urbanization rates substantially in excess of what is observed in 1700. Here urbanization rates are calculated by comparing the total implied population in cities with population of 5,000+ to total national population as in Acemoglu, Johnson, and Robinson (2005).

– they are observed at the local regional and national level in European history. Figure 11 shows how Zipf's Law emerged between 1400 and 1800 in the six leading national economies of Western Europe. Data in Russell (1972) on local urban systems in the high middle ages can be similarly used to show that Zipf's Law did not hold at the local level around 1200. Data from the Low Countries can also be used to show deviations from Zipf's Law at the local level in the mid-1400s (Stabel 2008).

Figure 11: The Emergence of Zipf's Law at the National Level



This graph examines city size distributions at the country level, using 20th century political boundaries to illustratively define national-level urban systems.

Table 8 compares data for cities in both databases. It shows that, on average, the sources give figures that are within 7 percentage points of each other. In keeping with the notion that measurement error increases as we reach back in the historical record, the deviations between the de Vries and Bairoch data decline over time: the correlation rises from 0.89 in 1500 to 1.00 in 1800; the ratio of recorded values approaches 1 and its standard deviation falls.

Given the deviations from Zipf's Law in the upper tail of the Bairoch data, it is natural to ask whether discrepancies are associated with city size. Figure 12 plots the de Vries data against the Bairoch data. It shows no evidence of systematic shortfalls in the populations that the Bairoch data record for large cities.[43] In the main body of

---

[43] Classical measurement error is not a plausible explanation for the observed deviations from Zipf's Law. See Gabaix (2008), who observes that: power laws are preserved under addition, multiplication and

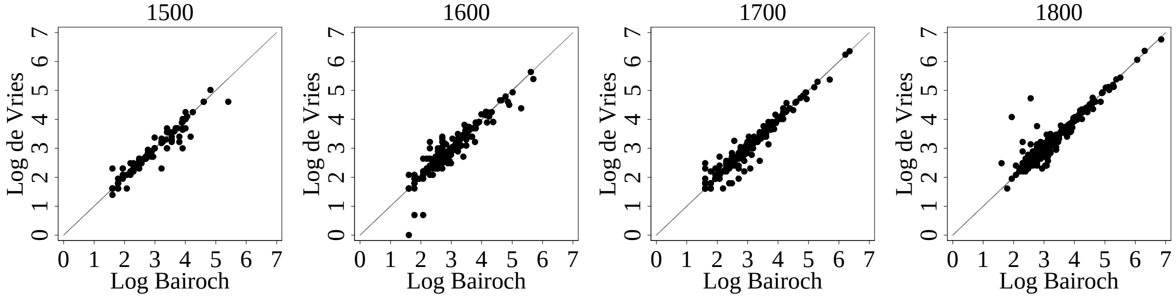
Table 8: Comparison of Source Data on City Populations

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
			Ratio of Bairoch Data to de Vries Data				
Year	Cities	Corr.	Mean	St. Dev.	Min.	Max.	Skewness
1500	117	0.88	1.07	0.30	0.50	2.50	2.92
1600	207	0.95	1.07	0.44	0.40	5.00	5.60
1700	250	0.99	1.02	0.22	0.42	2.31	2.83
1800	367	0.99	1.02	0.18	0.12	2.00	0.60

This table compares data on city populations from Bairoch, Batou, and Chèvre (1988) and de Vries (1984). Column (3) presents the correlation between recorded values. Columns (4) to (8) examine the ratio of these values.

the paper (Section 4.1), I show that the data would have to embody implausibly large non-classical measurement error for Zipf's Law to have actually held and (Section 5) that the observed deviations are consistent with the narrative evidence.

Figure 12: Comparison of Source Data on City Populations



Note: This figure plots city populations recorded in de Vries (1984) against corresponding values in Bairoch et al. (1988). The 45 degree line is shown to clarify where the Bairoch data provide larger (smaller) values.

C Small-Sample Estimators for Zipf Exponents

This appendix discusses the estimation of Zipf exponents and some properties of the Theil estimator.

Classically, Zipf's exponents have been estimated with standard OLS regressions of the form:

$$\ln R_i = \alpha - \beta \ln S_i + \epsilon_i \tag{5}$$

---

polynomial combination; multiplying by normal variables or adding non-fat tail noise does not change the exponent; and while noise will effect variances in empirical settings, it does not distort the exponent.

There are two problems with a standard OLS estimator. The first is that, even if the data generating process conforms strictly to a power law, the estimated coefficient $\hat{\beta}_{OLS}$ will be biased down in small samples. (As noted below, OLS standard errors are also biased down.) Gabaix and Ibragimov (2011) have proposed a remedy that reduces the bias in OLS coefficients to a leading order: adding a shift of -1/2 to the city rank data.

$$\ln(R_i - 1/2) = \alpha - \beta \ln S_i + \epsilon_i \tag{6}$$

For many applications this adjusted OLS approach may eliminate small sample bias.

The second problem with least squares is that any OLS estimator may be subject to gross errors in contexts marked by significant outliers due to the sensitivity of the OLS estimator to tail behavior. As He et al. (1990: 1196) note, "the tail performance of the least-squares estimator is found to be extremely poor in the case of heavy-tailed error distributions, or when leverage points are present in the design." Given the shape of the rank-size relation for European cities in the early modern era, this is a particular concern in the current setting.

The literature has discussed the Hill maximum likelihood estimator (MLE) as an alternative to OLS.[44] However, as Gabaix and Ioannides (2004) observe, the small sample biases associated with the Hill estimator can be quite high and very worrisome. Moreover, the Hill estimator is the MLE under the null hypothesis that the data generating process is a distributional (and specifically Pareto) power law, but is not the MLE if the empirical distribution is not Pareto. For these reasons, this paper does not present estimates using the Hill estimator.

Robust regression techniques have been designed for situations where sample sizes are small and/or outliers may have an undue impact on OLS estimates. A number of robust regression estimators use the framework provided by the median. In particular, the nonparametric estimator derived from Theil (1950) is intuitive, asymptotically unbiased, robust with small samples, allows us to go some distance in addressing the problem posed by outliers, and has not been studied in the Zipf's Law literature.[45] The Theil slope parameter is calculated as the median of the set of slopes that connect the complete set of pairwise combinations of the observed data points. Given observations (Y_k, x_k) for $k = 1, \ldots, n$, one computes the $N = n(n-1)/2$ sample slopes $S_{ij} = (Y_j - Y_i)/(x_j - x_i)$, $1 \leq$

---

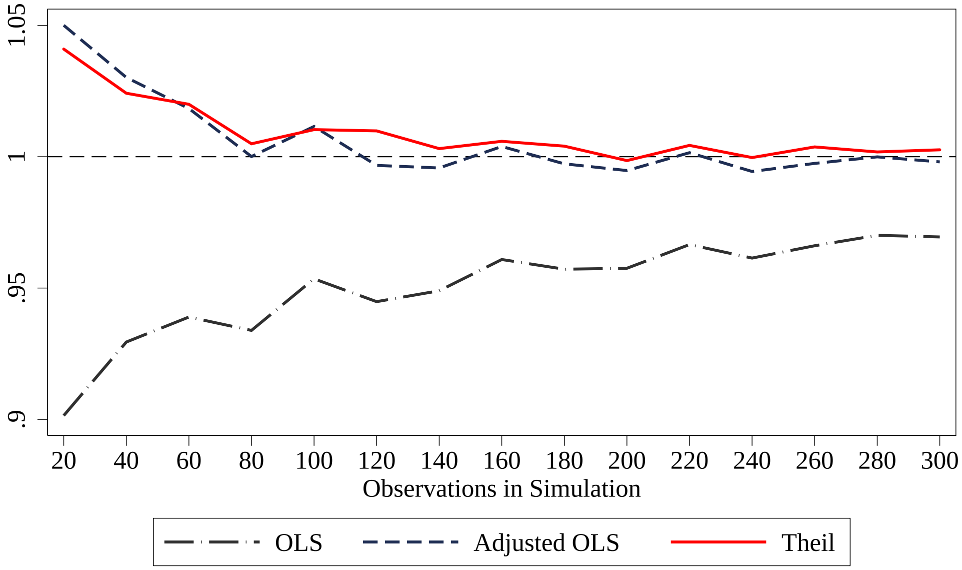
[44] For a sample of n cities with sizes S_i ordered so that $S_{(1)} \geq \ldots \geq S_{(n)}$, the Hill estimator is: $\hat{\beta}_H = (n-1)/\sum_{i=1}^{n} \left[\ln(S_{(i)}) - \ln(S_{(i+1)})\right]$.

[45] The repeated median regression suggested by Siegel (1982) and the least median of squares estimator suggested by Rousseeuw and Leroy (1987) are alternatives. But in the empirical context of this paper, they produce estimates that are virtually identical to the somewhat more elegant and parsimonious Theil (1950) estimator.

$i < j \leq n$. The Theil slope estimator is then: $\beta_T = \text{median}\{S_{ij}\}$. The corresponding constant term is: $\alpha_T = \text{median}_k\{Y_k - \beta_T x_k\}$. Hollander and Wolfe (1999) provide a generalization of the Theil estimator for cases where – as in the Bairoch data – the x_k are not all distinct.

The Theil estimator is competitive with the rank-adjusted OLS estimator suggested in Gabaix and Ibragimov (2011) in eliminating small sample bias. This is evident in Figure 13, which uses simulated data (generated by a process with Zipf exponent equal to 1) to compare small sample biases in estimated β's across OLS, rank-adjusted OLS, and Theil estimators.[46] Figure 13 reports mean estimates of the Zipf coefficient calculated over 1,000 simulations, each of which generates n synthetic observations from a distributional power law. To illustrate how estimates change with the sample size, Figure 13 reports the results as the number of observations in the simulations (n) rises from 20 to 300. While biased in small samples ($n < 80$), the small-sample bias in Theil estimates is relatively small. Moreover, the Theil estimate converges faster than OLS and as fast as the rank-adjusted OLS estimate.

Figure 13: Monte Carlo Estimates of the Zipf Exponent



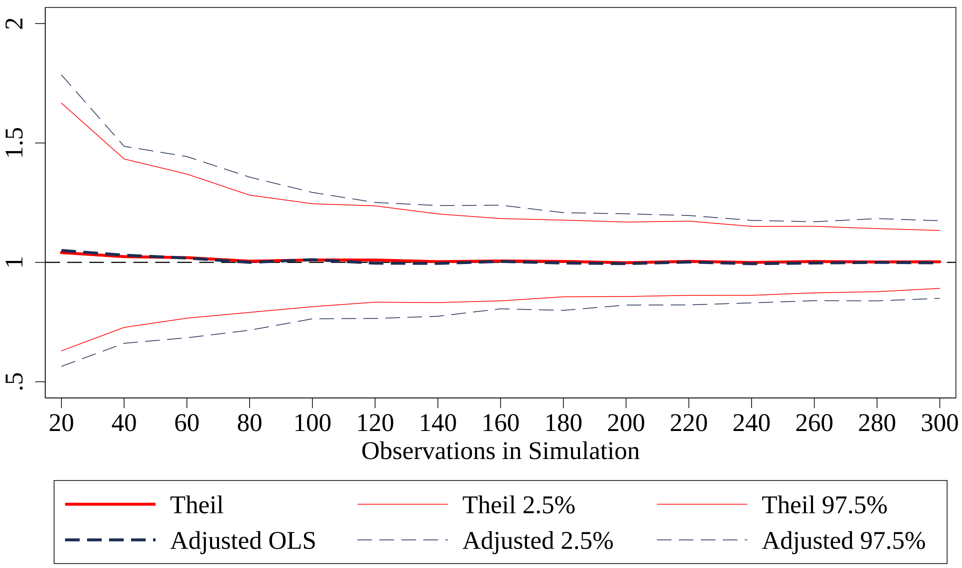
This figure presents the mean of Monte Carlo of the Zipf expoonent for synthetic data drawn from a power law with Zipf exponent of 1. The OLS estimates are from ordinary least squares regressions of log rank on log size from Equation (5). The "Adjusted OLS" regressions employ the Gabaix-Ibragimov adjustment and estimate Equation (6). The Theil estimates are for equation (5) using the Theil (1950) estimator. Mean estimates for each methodology are calculated over 1,000 simulations for each sample size, starting with samples of 20 observations and ending with samples of 300 observations.

---

[46] Data are constructed as follows. Sample n times from a uniform distribution on the unit interval to obtain x_i, $i = 1, \ldots, n$. Construct sizes $S_i = 1/x_i$ and rank the S_i's.

The Theil estimator also generates relatively precise estimates. Gabaix and Ibragimov (2011) show that, when we estimate power law exponents in small samples, OLS standard errors are biased down.[47] The confidence interval associated with Theil regression estimates similarly overstates the estimator's precision when data are drawn from a distributional power law.[48] To gauge and compare the true precision of these estimators, we can use Monte Carlo simulations. Figure 14 shows that the Theil estimates are more precise than the adjusted-OLS estimates. Future research may establish other empirical strategies, but Theil estimator effectively limits small sample bias as well as the estimators employed in the literature, while in addition being both robust to outliers and relatively precise.

Figure 14: The Precision of Estimates of Zipf Exponents



This figure presents the mean and bootstrapped 95 percent confidence interval associated with the Adjusted OLS (Gabaix and Ibragimov 2011) and Theil estimator. Mean estimates and confidence intervals for each methodology are calculated over 1,000 simulations for each sample size, starting with samples of 20 observations and ending with samples of 300 observations.

Given that the most widely used regression estimator is OLS, and that the Theil estimator is constructed as the median of the observed pairwise slopes, it is worth noting that OLS estimator is itself a weighted average of pairwise slopes. Using h to index the set of paired data points, define:

$$h \equiv (i,j) \quad X(h) \equiv \begin{bmatrix} 1 & x_i \\ 1 & x_j \end{bmatrix} \quad y(h) \equiv \begin{bmatrix} y_i \\ y_j \end{bmatrix} \quad b(h) \equiv X(h)^{-1} y(h)$$

---

[47] The true standard error of $\hat{\beta}$ in equation (6) is asymptotically $(2/n)^{0.5}\hat{\beta}$.

[48] See Hollander and Wolfe (1999) for calculation of confidence intervals on Theil slope parameter.

Under this notation, the OLS estimator is: $\beta_{OLS} = \sum_{h=1}^{N} w(h)b(h)$, where the weights are defined as: $w(h) = |X(h)|^2 / \sum_{h=1}^{N} |X(h)|^2$. These weights are proportional to the distance between design points. As Koenker (2005: 4) observes this is a fact that, "in itself, portends the fragility of least squares to outliers."

D OLS Regression Tests for Power Laws

This section presents both OLS and Gabaix (2009) estimates used to test for power laws.

Indexing cities with i and denoting city size S and city rank R, Zipf's exponents have classically been estimated with OLS regressions of the form:

$$\ln R_i = \alpha - \beta \ln S_i + \epsilon_i \tag{7}$$

A number of studies suggest employing a regression augmented with a quadratic term to detect non-linearities and deviations from distributional power laws[49]:

$$\ln R_i = \beta_0 - \beta_1 \ln S_i + \beta_2 (\ln S_i)^2 + \nu_i \tag{8}$$

As discussed below, the standard errors associated with this model are biased down. However, I present historical estimates of equation (8) to facilitate comparison with existing studies using non-historical data. Table 9 shows that between 1500 and 1700, and certainly by 1800, a "modern" city size distribution emerged in Western Europe. In contemporary data on a large sample of countries, Soo (2005) finds estimates of Zipf exponents ranging from 0.7 to 1.5. From 1700, Western European cities have a Zipf exponent $\hat{\beta}_1 \in (0.7, 1.5)$ and modest non-linearity in the logarithmic rank-size relation: $\hat{\beta}_2$ is "small" and by 1800 vanishes.

However, the estimates in Table 9 should be treated with caution. It can be shown using synthetic data from a pure power law distribution that heteroskedasticity-robust standard errors associated with equation (8) exhibit downward bias in finite samples.[50] It follows that the statistical significance of $\hat{\beta}_2$ is not a robust criterion on which to base rejection of Zipf's Law. Hence Table 9 should be read as indicating the existence (or absence) of gross departures from Zipf's Law, not as a precise test.

As described in the main text, the alternate test developed in Gabaix (2009) relies on

---

[49] As Soo (2005) notes, this regression may be viewed as a weak form of the Ramsey RESET test.

[50] Ranking induces a positive correlation between residuals which escapes conventional estimation. See Gabaix and Ioannides (2004: 2348).

Table 9: OLS Regression Analysis of Deviations from Zipf's Law

	(1)	(2)	(3)	(4)	(5)	(6)
	Outcome: Log Rank					
	1300	1400	1500	1600	1700	1800
Ln Population	0.30***	0.13	-0.20*	-0.82***	-0.95***	-1.36***
	(0.08)	(0.25)	(0.10)	(0.05)	(0.05)	(0.05)
Ln Population Squared	-0.28***	-0.22***	-0.20***	-0.08***	-0.04***	-0.00
	(0.02)	(0.05)	(0.02)	(0.01)	(0.01)	(0.01)
Observations	255	187	321	514	539	1311
R^2	0.994	0.986	0.993	0.997	0.997	0.997

The estimated regression is: $\ln R_i = \beta_0 - \beta_1 \ln S_i + \beta_2 (\ln S_i)^2 + \nu_i$, where R_i is city rank and S_i is city population. Each column presents estimates for the year indicated. Bootstrapped standard errors in parentheses. Statistical significance denoted * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

an OLS regression:

$$\ln(R_i - 1/2) = \delta_0 + \delta_1 \ln S_i + \delta_2 (\ln S_i - S^*)^2 + \epsilon_i \tag{9}$$

where $S^* \equiv \text{cov}[(\ln S_i)^2, \ln S_i]/2\text{var}[\ln S_i]$ and the shift of -1/2 provides the optimal reduction in small sample bias in the OLS setting.[51] Under the Gabaix test, we reject the null hypothesis of a power law with 95 percent confidence if and only if $|\hat{\delta_2}/\hat{\delta_1}^2| > 1.95(2n)^{-0.5}$. Table 10 shows that implementing the Gabaix (2009) test that we can reject Zipf's Law in Western Europe through 1600, but not from 1700 forwards (see also Table 3 in the main text).

E Additional Historical Evidence

The main text presents discussion on the historical relationship between agricultural productivity and city populations. Figure 15 shows how urbanization rates were correlated with estimates of agricultural total factor productivity in nine macroeconomies, using evidence on agricultural productivity from Allen (2003).

The main text also discusses the magnitude of grain imports into the Netherlands in the early modern era. Figure 11 provides evidence on the magnitude of these imports in the mid-1500s, the 1590s, and the mid-1600s. To benchmark these figures, in 1600 the

---

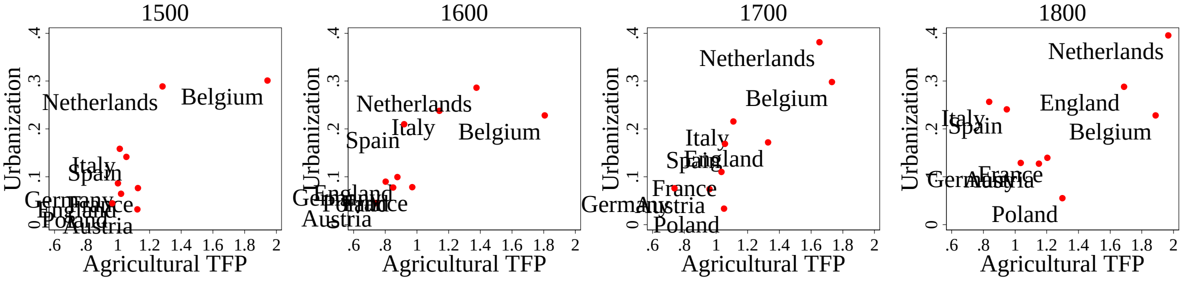
[51] An earlier literature examined Zipf's Law with regressions: $\ln(R_i) = \beta_0 + \beta_1 \ln S_i + \beta_2 \ln S_i^2 + \nu_i$. As discussed in Gabaix (2009), heteroskedasticity-robust standard errors will be biased down in this specification and the statistical significance of $\hat{\beta}_2$ is not a robust criterion for a test of Zipf's Law. However, to facilitate comparison with existing studies, Appendix D presents results from this specification which support the conclusion that Zipf's Law emerged in Western Europe 1500-1800.

Table 10: Gabaix-Ibragimov Test for Power Law

	(1)	(2)	(3)	(4)	(5)	(6)
	Outcome: Adjusted Log Rank					
	1300	1400	1500	1600	1700	1800
Parameter δ_1	-1.31***	-1.15***	-1.35***	-1.33***	-1.22***	-1.40***
	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.00)
Parameter δ_2	-0.33***	-0.26***	-0.24***	-0.11***	-0.06***	-0.02***
	(0.03)	(0.03)	(0.01)	(0.01)	(0.01)	(0.00)
Observations	255	187	321	514	539	1311
R^2	0.991	0.989	0.994	0.997	0.998	0.997
Reject Power Law	Yes	Yes	Yes	Yes		

A regression-based test of deviations from Zipf's Law. Each column presents regression estimates for the year indicated. The estimated model is: $\ln(R_i - 1/2) = \delta_0 + \delta_1 \ln S_i + \delta_2 (\ln S_i - S^*)^2 + \epsilon_i$, where R_i is city rank, S_i is city population, and $S^* \equiv \text{cov}[(\ln S_i)^2, \ln S_i]/2\text{var}[\ln S_i]$. We reject the null hypothesis of a power law with 95 percent confidence if and only if $|\hat{\delta_2}/(\hat{\delta_1})^2| > 1.95(2n)^{-0.5}$. Standard errors adjusted to correct for the positive autocorrelation of residuals induced by ranking. Statistical significance denoted * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure 15: Agricultural TFP and Urbanization Across European Economies



This graph plots agricultural TFP estimates from Allen (2003) against urbanization rates from Acemoglu et al. (2005).

Bairoch data provide city population data for 27 Dutch cities with a total population of 429 thousand inhabitants.

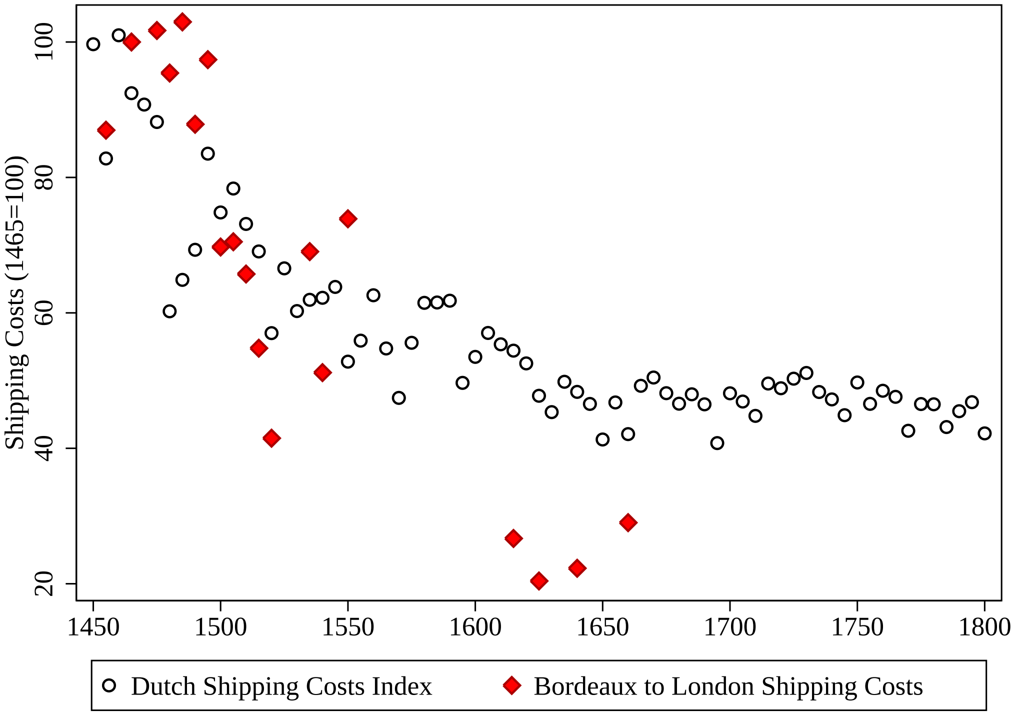
Table 11: Dutch Imports of Baltic Grain

Period	Hectolitres per Year	Enough to Feed
1550s	1,053,500	263,375
1590s	1,745,800	436,450
1640s	2,859,500	714,875

Data from van Tielhof (2002). Calculations assume consumption of 4 hectolitres per person per year. In 1600, the 24 largest Dutch cities had 421 thousand inhabitants.

Figure 16 presents evidence on the cost of Dutch shipping and the cost of shipping wine from Bordeaux to London, from van Tielhof and van Zanden (2009) and Menard (1991), respectively. These data support the calculations and discussion presented in the main text.

Figure 16: Real Maritime Transport Costs



This graph plots measures of real maritime transport costs. The Dutch shipping cost index is from van Tielhof and van Zanden (2009). Nominal Dutch shipping costs are deflated by the wholesale price index also from van Tielhof and van Zanden (2009). Bordeaux-London shipping costs are from Menard (1991) and describe the costs associated with transporting wine, deflated by the English consumer price index also from Menard (1991).

E.1 Stability and Dynamics in City Size Hierarchies

If the distribution of city populations is determined by locational fundamentals, the fact that Zipf's Law emerged over time suggests that the fundamentals may be dynamic.[52] If geographic advantages are not fixed, it is natural to wonder whether the rank hierarchy of city populations is persistent: Are certain locations persistently the most populous?

Two notable findings have provided support for a geographic fundamentals view of the distribution of economic activity. The first is Davis and Weinstein's (2002) observation that a *regional* analogue of Zipf's Law holds over many centuries in the Japanese data

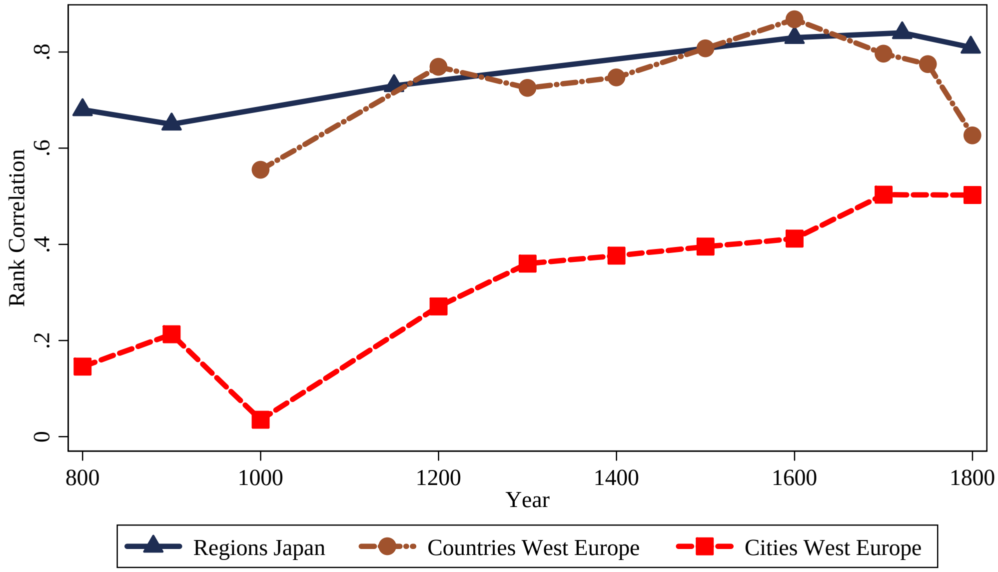
---

[52] The period and processes studied in this paper thus differ from the setting studied by Bleakley and Lin (2012), which examines North American data and documents how agglomerations at locations which enjoyed natural advantages due to the geography of waterborne transport in the pre-rail road era maintained persistent development advantages through the 20th century – after transportation technology changed and their initial location-specific geographic characteristics ceased to be economically salient.

and that Japanese regional population densities in the past are highly correlated with contemporary population densities. The second is that 20th century city size hierarchies have been stable in the face of large temporary shocks.[53]

In contrast, long run evidence suggests European city size hierarchies were relatively dynamic. Figure 17 shows that the high correlations Davis and Weinstein find between historical and contemporary populations in regional data from Japan are matched in the European data on *national* population densities, but not in the European data on *city* populations, which have been less persistent. This suggests that units matter and that a distinction between regional and city-level population data may be useful.

Figure 17: Correlations between Historic and Contemporary Populations



This figure plots the correlations between historic and contemporary populations. Correlations between historic and contemporary populations for Japanese regions are from Davis and Weinstein (2002). Correlations for the populations of countries in Western Europe rely on contemporary national population data for the year 2000 from Eurostat (2009) and Statistics UK (2001) and period-by-period historic national population data from Acemoglu, Johnson, and Robinson (2005). Correlations for cities in Western Europe rely on contemporary city population data are Brinkhoff (2008) and historic data from Bairoch, Batou, and Chèvre (1988). Rank correlations for regional and national data are calculated using population densities (i.e population divided by land area in square kilometers).

The data similarly reveal considerable churning in European city hierarchies between 1300 and 1800. Table 12 presents evidence on city transitions across the population distribution between 1300 and 1800. It shows that 56% of cities in the upper quartile of the size distribution in 1300 remained there in 1800 – implying that 44% dropped into lower quartiles, including 5% which dropped to the smallest quartile. Cities that

---

[53] For example, Davis and Weinstein (2002) and Brakman et al. (2004) study the bombing of German and Japanese cities during the 1940s as quasi-natural experiments and find these shocks had little long run impact on the distribution of city populations in Japan and Germany.

transitioned from the largest to the smallest quartile include Leon, Narbonne, and Speyer. Cities which dropped from the largest to the second-from-bottom quartile include Aquila, Burgos, Perugia, Tours, and Ypres.

Table 12: City Size Transition Matrix 1300 to 1800

	Population Quartile in 1800			
Quartile in 1300	Q1: 1%-25%	Q2: 25%-50%	Q3: 50%-75%	Q4: 75%-100%
Q1: 1%-25%	41.18	30.88	20.59	7.35
Q2: 25%-50%	30.91	30.91	20.00	18.18
Q3: 50%-75%	20.97	30.65	27.42	20.97
Q4: 75%-100%	5.08	8.47	30.51	55.93

This table presents the transition matrix describing the probability of transitioning between city population quartiles in 1300 and 1800 for 244 cities with population observed in both periods. In 1300, Q1 comprised cities with 5-7 thousand inhabitants; Q2 8-11 thousand; Q3 12-20 thousand; and Q4 21 thousand or more. In 1800, Q1 comprised cities with 5-9 thousand; Q2 10-16 thousand; Q3 17-30 thousand; and Q4 31 thousand or more.

More broadly, in the European data we observe notable churning in the identity of the very largest cities over the course of centuries. It is not just that in 1500 the largest cities were concentrated in Southern Europe, while in 1800 the largest cities were concentrated in Northwestern Europe. There were also sharp shifts in urban populations at local levels. Cologne was the largest German city between 1200 and 1500; today it is the 7th largest. Augsburg went from being the largest German city in 1600 to 8th largest in 1800 and 24th in 2006. In 1400, Madrid was a village while Cordoba and Granada had populations of 60 and 150 thousand. In 1800, Madrid had a population of 160 thousand, where Cordoba and Granada had populations of 40 and 70 thousand. In 1000 AD, Laon was the largest city in France with a population of 25 thousand, while Caen, Tours, Lyon, and Paris all had approximately 20 thousand inhabitants. In 2006, Laon had 27 thousand inhabitants, Caen had 186 thousand, Tours 307 thousand, Lyon 1.4 million, and Paris over 10 million. Ostia (population 50 thousand in the 2nd century), Pozzuoli (65 thousand in the 2nd century), and Brindisi were great port cities in the Roman era, but fell into disuse and remained small population centers over the early modern era. In 200 AD Rome was Europe's largest city with a population of nearly one million. Between 800 and 900 AD, Rome had a population of approximately 50 thousand and was Western Europe's second largest city. In 1300, Rome was the 32nd largest city in Western Europe. Between 1500 and 1800, Rome was among the 10 largest cities. By 1850 it was 17th.[54]

---

[54] For historical populations see Bairoch, Batou, and Chèvre (1988), Meiggs (1973), and Stillwell (1976). Contemporary French populations are for urban agglomerations and are from Brinkhoff (2008).

F Model

F.1 Motivation

Leading theories explain Zipf's Law as the outcome of a random growth process. Rossi-Hansberg and Wright (2007) have shown that the slight curvature observed in log rank-log size plots of contemporary city population data may reflect a negative correlation between city sizes and city growth rates. Intuitively, this curvature emerges when small cities tend to grow quickly and "escape" to become mid-sized, and when larger cities tend to grow slowly, leaving the largest cities smaller in size and the small cities fewer in number than they "should be." A similar, but more pronounced curvature characterizes the historical data. As shown below, this curvature is observed when and where growth rates were negatively correlated with city size over long periods.

I incorporate Rossi-Hansberg and Wright's insight in a simple model of city growth.[55] The model contains a feature that may deliver non-random growth: land may be a fixed argument in production, generating decreasing returns to scale. When this feature is "shut off," the model reduces to the random growth model in Gabaix (1999b). This provides a very stark description of the environment, in which the decline of transport costs, increases in agricultural productivity, and innovations in activities that do not use land all work to eliminate the land argument in city production.[56]

F.2 Environment

The model has overlapping generations. At any time t, cities indexed with i have old residents N_{it}^o and young residents N_{it}^y, with old people dying at some rate δ. The overlapping generations structure has a first period in which potential workers are born young and decide if and where to migrate (paying some fixed migration cost x). In subsequent periods workers are old and live out their days without further migration.[57]

There are city-specific amenity shocks a_{it} due to some combination of policy and nature. In particular:

$$a_{it} = \epsilon_{it}(1 - \tau_{it}) \tag{10}$$

---

[55] In Rossi-Hansberg and Wright's model, industrial specialization accounts for urban hierarchies. However, Zipf's Law emerged when industrial and functional specialization was very limited, suggesting that another mechanism may have been at work.

[56] Slightly more complicated models could examine the evolution of the land constraint and trade offs between increasing returns and transport/congestion costs.

[57] Workers are young once and typically old for multiple periods.

ϵ_{it} is an iid city-specific shock and $\tau_{it} \in (0,1)$ is a city-specific distortion. Without loss of generality, the amenity shocks a_{it} enter utility multiplicatively:

$$u(c) = a_{it}c \tag{11}$$

Production is Cobb-Douglas in technology (A), labor (young N^y and old N^o), and land (L)[58]:

$$Y_{it} = A_{it}(N_{it}^y)^{\alpha}(N_{it}^o)^{\beta}(L_{it})^{1-\alpha-\beta} \tag{12}$$

Assume that $\alpha, \beta \in (0,1)$ and that $\alpha + \beta \leq 1$. Where $\alpha + \beta = 1$, production is CRS in labor. By assumption, city residents own labor but not land.[59] The wage is the marginal product of labor and is consumed in each period:

$$c_{it} = w_{it} = \frac{\partial Y_{it}}{\partial N_{it}^j} \quad j \in \{y, o\} \tag{13}$$

The aggregate number of young potential migrants is determined by a "birth rate" n_t and the total number of mature agents. The birth rate can equally be taken as a description of the migration rate from the non-urban sector. The number of young agents arriving in each city is endogenous.[60]

F.3 Analysis of City Growth – The General Case

Young people choose a city i subject to city-specific migration taxes τ_{it} and given the existing distribution of populations (wages). The individual maximization problem reduces to[61]:

$$\max_i \, a_{it}w_{it}$$

---

[58] Production is modelled without a capital argument in the interest of parsimony. NB: In the pre- and early modern era, fixed capital was important in the rural economy but less critical in the cities. See Cipolla (1982).

[59] This assumption raises the question: who receives rents on urban land? One can assume following Henderson (1974, 2005) that each city is owned by a single private land developer. This was the situation in many Eastern European cities, which were owned by feudal lords. Alternately, assuming that a (small) class of urban landowners receive and consume the marginal product of urban land, would not change the basic story. Historically, the evolution of city populations was largely driven by the evolution of non-landowning populations. In the interest of parsimony, the model focuses on these agents.

[60] As in Gabaix (1999b), the model here pursues parsimony. To that end, it abstracts from interesting questions concerning the interaction between agglomeration economies and congestion costs.

[61] For simplicity, agents make a calculation based on utility in the current period, completely discounting future periods (and potential tax changes). A tax on wages leads to the same results.

In equilibrium with free mobility: $u_{it} = u_t$. It follows that:

$$w_{it} = \frac{u_t}{a_{it}} \tag{14}$$

Because young people earn wages equal to their marginal product, and wages equalize across age groups, we have that:

$$w_{it} = \alpha A_{it}(N_{it}^y)^{\alpha-1}(N_{it}^o)^{\beta}(L_{it})^{1-\alpha-\beta} \tag{15}$$

Combining (10), (14), and (15), we get an expression for the number of new-comers in the representative city:

$$N_{it}^y = (N_{it}^o)^{\frac{\beta}{1-\alpha}}(A_{it})^{\frac{1}{1-\alpha}}(L_{it})^{\frac{1-\alpha-\beta}{1-\alpha}}(1-\tau_{it})^{\frac{1}{1-\alpha}}\left(\frac{\alpha\epsilon_{it}}{u_t}\right)^{\frac{1}{1-\alpha}} \tag{16}$$

The representative city growth rate is:

$$g_{it}^N \equiv \frac{\Delta N_{it}}{N_{it}} = \frac{N_{it}^y - \delta N_{it}^o}{N_{it}^o} \tag{17}$$

Substituting with equation (16) gives:

$$g_{it}^N = (N_{it}^o)^{\frac{\beta+\alpha-1}{1-\alpha}}(A_{it})^{\frac{1}{1-\alpha}}(L_{it})^{\frac{1-\alpha-\beta}{1-\alpha}}(1-\tau_{it})^{\frac{1}{1-\alpha}}\left(\frac{\alpha\epsilon_{it}}{u_t}\right)^{\frac{1}{1-\alpha}} - \delta \tag{18}$$

A distortion hitting productivity would have an identical growth rate impact.

F.4 Case 1: Random Growth with No Distortions

The conventional argument in the Zipf's Law literature is that growth rates are independent of city size. This argument typically embodies three assumptions: fixed factors are not important in urban production; productivity does not vary with population across cities; and distortions are independent of city size. When land does not enter production $\alpha + \beta = 1$. The idea that productivity and distortions (e.g. migration costs) do not vary with city size can be captured by assuming: $\tau_{it} = \tau_t$ and $A_{it} = A_t$. To consider the case without distortions let $\tau_t = 0$. Substituting into equation (18) gives:

$$g_{it}^N = (A_t)^{\frac{1}{1-\alpha}}\left(\frac{\alpha\epsilon_{it}}{u_t}\right)^{\frac{1}{1-\alpha}} - \delta \tag{19}$$

Since the only city-specific argument on the right-hand side of (19) is the iid random shock ϵ_{it}, the rate of population growth is independent of city size. Provided we have some (arbitrarily small) reflecting barrier that keeps cities from getting "too small," random growth delivers Zipf's Law. This is the model in Gabaix (1999b).[62]

F.5 Case 2: Non-Random Growth Due to Fixed Land

Assume that migration costs are constant across cities, but land has some positive income share. For simplicity, normalize $L_{it} = L_i = 1$ and assume that $A_{it} = A_t$. Assume also no distortions: $\tau_{it} = 0$. We now have the following variant of equation (18):

$$g_{it}^N = (N_{it}^o)^{\frac{\beta+\alpha-1}{1-\alpha}} (A_t)^{\frac{1}{1-\alpha}} \left(\frac{\alpha \epsilon_{it}}{u_t} \right)^{\frac{1}{1-\alpha}} - \delta \tag{20}$$

Here land has a positive income share because $\alpha + \beta < 1$. This fact secures the key feature of (20): city growth rates decline in population when land is fixed.

Broadly, one can view the long pre-modern era as one in which land entered city production and land was more or less fixed. Under a fixed-land regime, growth rates are negatively correlated with city populations. Small cities will tend to draw high growth rates and become mid-sized. Similarly, big cities will tend to draw low growth rates and remain relatively small. Thus a fixed factor can deliver a distribution of growth rates in keeping with the curvature we observe in European city size distributions between 1300 and 1600.

---

[62] Gabaix assumes technology is fixed ($A_t = 1$).