Cities, Markets, and Growth: The Emergence of Zipf's Law

Jeremiah Dittmar*

August 10, 2011

Abstract

Zipf's Law characterizes city populations as obeying a distributional power law and is supposedly one of the most robust regularities in economics. This paper documents, to the contrary, that Zipf's Law only emerged in Europe 1500-1800. Until 1500, land entered city production as a quasi-fixed factor. Big cities grew relatively slowly and were "too small." After 1500, developments in trade, agricultural productivity, and knowledge-based activities relaxed this constraint. As a result, city growth became size independent and Zipf's Law emerged. This urban transformation occurred in the centuries immediately preceding the industrial revolution and the onset of modern economic growth.

Keywords: Cities, Growth, Zipf's Law, Power Laws.

JEL Classification: N13, N33, N93, O11, O18, R11, R12

*Email: dittmar.academic@gmail.com. Mail: Institute for Advanced Study, School of Social Science, Einstein Drive, Princeton, NJ 08540. I thank Barry Eichengreen, Chad Jones, Christina Romer, Brad DeLong, Greg Clark, Suresh Naidu, and seminar participants at UC Berkeley, American, UC Davis, Queens (CUNY), Dartmouth, UBC, and Stanford for comments, discussions, and suggestions. The errors are mine.

1 Introduction

Economists have identified an underlying order in urban hierarchies. Zipf's Law characterizes city populations as obeying a distributional power law (a Pareto distribution) and is supposedly one of the most robust regularities in all of economics. Krugman (1996a: 39) observes that this distributional regularity is "suspiciously like a universal law." Gabaix (1999a: 129) notes that it appears to hold in all economies and periods for which there are data. This paper shows, to the contrary, that Zipf's Law only emerged in Western Europe between 1500 and 1800. It documents how Zipf's Law emerged with the development of markets in relatively advanced economies in the centuries immediately preceding the onset of modern, capitalist economic growth.

The leading theories tie Zipf's Law to either (1) random growth or (2) the underlying distribution of geographic advantages or "locational fundamentals" (Krugman 1996a; Davis and Weinstein 2002; Gabaix 2008). The random growth theory provides a bench-mark for thinking about city population dynamics. The locational fundamentals theory carries important predictions about the way geography shapes economic life. Zipf's Law is thus more than a mathematical curiousity. Zipf's Law is a key stylized fact and an important constraint on the set of admissable models of urban growth (Gabaix 1999a). Examining its emergence leads us to ask important questions about cities and the determinants of growth broadly speaking.

This paper exploits data on the populations of European cities since 800 AD to test the random growth and locational fundamentals theories. I document that Zipf's Law did not hold before 1500 and only emerged after city growth became random. Similarly, I find significant churning in the distribution of European city populations – evidence that while geography matters, it is not destiny. These quantitative findings are supported by a rich body of evidence historians have assembled to characterize the historical 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.

In addition to testing the random growth and locational fundamentals theories, I document that significant deviations from Zipf's Law reflected big impediments to trade and limits on the operation of markets. Historically, a land constraint limited the growth of big cities, which grew relatively slowly and were far smaller than Zipf's Law would lead us to expect until at least 1500. After 1500, developments in trade, rising agricultural productivity, and the sharp growth of knowledge-based activities relaxed the land constraint –

making it 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. This paper thus documents how a recognizably modern city system developed with commercial activity in Europe in the centuries immediately preceding the Industrial Revolution.

2 Literature

2.1 Zipf's Law

Zipf's Law for cities can be characterized in two ways.[1] 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[2]:

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

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

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

In the literature, 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 takes an agnostic position on the range of acceptable β's.[3]

[1] The proper entities are urban agglomerations, which are what this paper analyzes.

[2] Even if the data generating process conforms to equation (1), equation (2) only holds approximately. Gabaix (1999b, 2008) provides discussion and derivations.

[3] 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. Soo (2005) finds $\beta \in [0.7, 1.4]$ across 75 contemporary economies.

2.2 Theories

Three types of theories have been advanced to explain Zipf's Law. Random growth theories explain Zipf's Law as the outcome of a growth process in which all cities – big and small – draw growth rates from some common distribution. Geographic theories explain Zipf's Law as reflecting the distribution of natural advantages across locations. Static but non-geographic theories explain Zipf's Law as an equilibrium outcome given agglomeration economies, congestion costs, and a nearly Pareto distribution of talent.

Gabaix (1999b, 2008) 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. Beyond random growth, the key assumption in Gabaix (1999b) is that there is an arbitrarily small reflecting barrier that prevents cities from getting "too small."[4] Recent theoretical work has further explored how random growth may deliver Zipf's Law. Cordoba (2004) 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.[5] In Cordoba (2004) 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.

Against theories that center on random growth, Krugman has suggested a geographic explanation. Krugman (1996b) observes that the physical landscape is not homogeneous, and that the distribution of propitious locations may follow a power law and thus account for the size distribution of cities.[6] Davis and Weinstein (2002: 1269-1270) similarly argue that city size hierarchies are determined by locational fundamentals that are essentially fixed over time. Davis and Weinstein 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." In their view, it is not city growth, but the fundamental

[4]This assumption is consistent with the historical evidence. 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) and are reviewed in Gabaix (1999b, 2008).

[5]In Rossi-Hansberg and Wright (2007) Zipf's Law 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.

[6]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.

economic characteristics of locations that are random.

Geographic theories do not have a strict monopoly on static models delivering Zipf's Law. Behrens, Duranton, and Robert-Nicoud (2010) develop a static model of cities in which talent follows an (approximate) power law. In their model, the fixed distribution of talent, agglomeration economies, and congestion costs together deliver Zipf's Law.

2.3 Evidence

In recent work, Ioannides and Overman (2004) show that contemporary city growth in the USA appears to be random. But 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, Overman, Rossi-Hansberg, and Schmidheiny (2007).[7] 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. They argue that this results when small cities grow quickly and large cities grow slowly. I return to this point below.

Davis and Weinstein (2002) provide evidence in support of the locational fundamentals view. Davis and Weinstein find that in Japan a regional analogue to Zipf's Law held across time periods stretching back thousands of years and that the hierarchy of regional population densities in Japan has been relatively stable over many centuries. They 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 and that random growth theories are flawed.

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.[8] Russell (1972) provides data revealing that, from the perspective of Zipf's Law, the largest cities in the urban systems of medieval Europe were relatively small. Stabel (2008) provides data that document similar deviations from Zipf's Law in the Low Countries in 1450 at both the aggregate and the provincial level. Archaelogical data also confirm departures from Zipf's Law across a

[7] A debate 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, Rosenfeld, Rybski, Gabaix, and Makse (2010) find that when cities are defined according to economic rather than legal or administrative criteria, Zipf's Law holds for agglomerations with populations as small as 12,000 in the USA and 5,000 in the UK.

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

range of pre-modern or non-capitalist economies (Johnson 1980; Drennan and Peterson 2004; and Savage 1997).[9] Thus de Vries (1984; 1990: 52) observes that urban systems may not always conform to Zipf's Law and 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."[10]

3 Data

In this section I present 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 paper employs data on European city populations from Bairoch, Batou, and Chèvre (1988). Their approach is to identify the set of cities that ever reached 5,000 inhabitants between 1000 and 1800, and then to search for population data for these cities in all periods. The data record (in thousands) the populations of urban agglomerations, not simply populations within administratively defined boundaries.[11] These data – henceforth the "Bairoch data" – are recorded every 100 years 800-1700 (except 1100) and every 50 years 1700-1850.

This paper only examines cities with population of at least 5,000. It further restricts the principal analysis to 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 1 shows the locations of the historic Western European cities examined in this paper.

Below I test for measurement error in several ways and document that measurement error cannot account for the observed deviations from Zipf's Law. I compare the Bairoch data to the most comprehensive independent source for city population data, the database

[9]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.

[10]In de Vries (1984), analysis is restricted cities with population of 10,000 or more and the period 1500-1800. This paper examines a panel of cities with population of 5,000 or more over the period 800-1800.

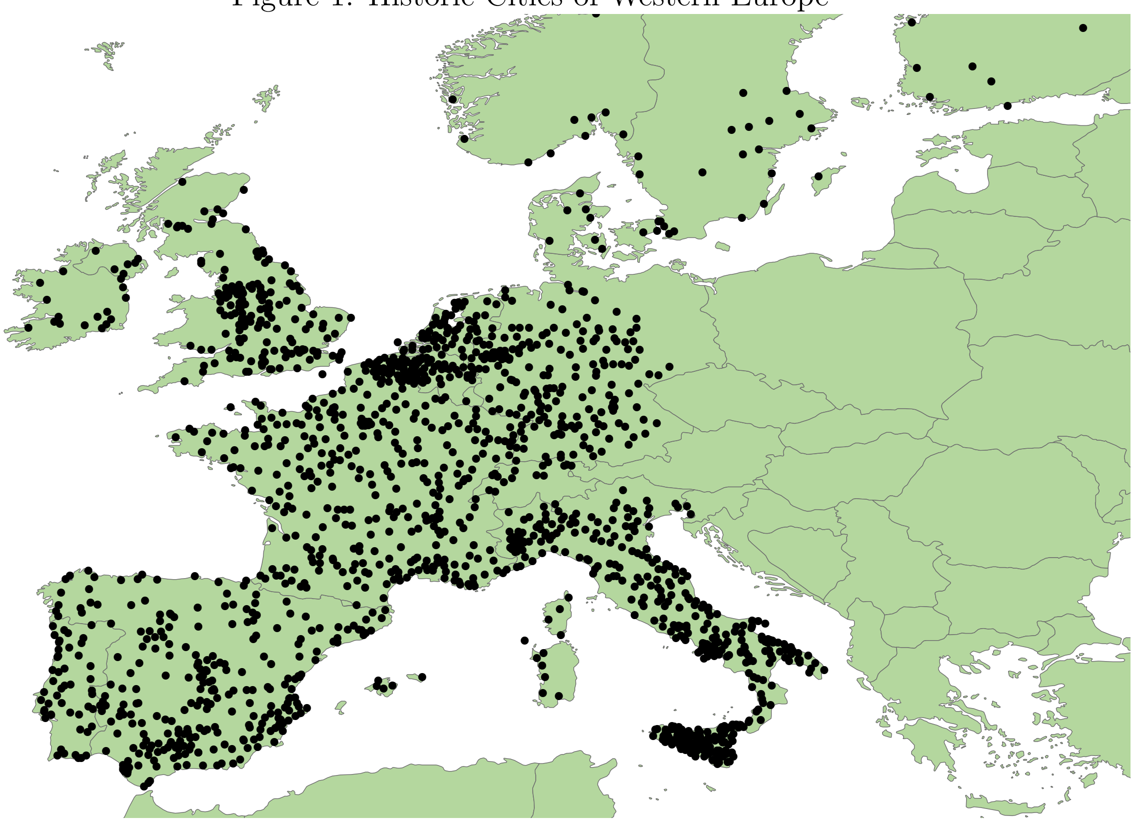
[11]Bairoch, Batou, and Chèvre (1988: 289) make a special effort to include, "the 'fauborgs', the 'suburbs', 'communes', 'hamlets', 'quarters', etc. that are directly adjacent" to historic city centers. Bairoch et al. draw data from primary and secondary sources. Prior to publication the data were reviewed by 6 research institutes and 31 regional specialists in urban history. The fact that populations are recorded in thousands, and not as continuous counts of individuals, implies the presence of measurement error which I discuss below.

Table 1: City Populations and City Growth in Western Europe

	Population at Beginning of Period			Annualized Population Growth		
Period	Cities	Mean	St. Dev.	Cities	Mean	St. Dev.
(1)	(2)	(3)	(4)	(5)	(6)	(7)
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%

Note: This table records the number of 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.

Figure 1: Historic Cities of Western Europe



in de Vries (1984).[12] I document that there is no evidence of systematic shortfalls in the populations that the Bairoch data record for large cities (see Appendix B).[13] In section 4.2, I show that the data would have to embody implausibly large non-classical measurement error for Zipf's Law to have held: in 1500 cities like Paris and Naples would have required populations about three times as large as observed and as large as their populations in 1800. Similarly, if missing or mismeasured data on small cities were to account for the observed deviations from Zipf's Law, this would imply counterfactually high urbanization rates in periods with big deviations from Zipf's Law. Finally, in section 5, I show that the observed deviations from Zipf's Law are consistent with a rich body of evidence on the demography of historic cities.

3.2 Regional Classification of Cities

This section explains why it makes sense to examine the Zipf's Law in Western Europe as a whole and how differences in the institutional environment distinguished cities in Western Europe from cities in Eastern and Ottoman Europe.[14]

The unit of analysis in contemporary research on Zipf's Law is typically the national economy. However, a transnational perspective is appropriate for an analysis of city growth in European history. Between 1000 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; Landes 1998; Jones 1981; and Bosker, Buringh, and van Zanden 2008). 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 1500-1800 was mirrored by its emergence over the same period at the local and national level. These facts are documented in Appendix C.

The distinction between Western and non-Western cities was determined by two key aspects of economic institutions. The first was the presence of institutions securing mu-

[12]The Bairoch data cover 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) cover cities that reached a population of 10,000 between 1500 and 1800, record populations for 379 cities, and cover only periods 1500-1800.

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

[14]An earlier version of this paper (Dittmar 2010) examines cities in Eastern and Western Europe – and documents the impact institutional restrictions on labor mobility had on city growth and the emergence of Zipf's Law in Eastern Europe.

nicipal autonomy for cities. The second was the nature of the institutions determining the possibilities for mobility between the rural and urban sectors.

Cities in Western Europe developed in a distinct institutional environment. Town charters in the West guaranteed townspeople the right to legal proceedings in town courts, the right to sell their homes and move, and freedom from obligations associated with serfdom (e.g. arbitrary taxation, the provision of labor services, and most forms of military service). These institutions fostered geographic mobility, relatively secure property rights, and the growth of urban commerce.[15]

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.[16] An extensive literature documents the importance of the Elbe River (which cuts through Eastern Germany) as an institutional boundary distinguishing Western Europe from a central Eastern Europe in which the legal institutions of serfdom were strengthened after 1500 and city growth was subsequently distorted.[17] 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.[18]

In light of these facts, this paper examines the distribution and dynamics of Western European city populations. Consistent with the institutional distinction identified in the historical literature, I class as "Western European" cities located West of the Elbe River and/or its tributary the Saale and outside Ottoman Europe. See Map 1 above.

[15]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).

[16]See, for instance, Bideleux and Jeffries (2007), Süchs (1988), Kamiński (1975), Topolski (1982), and Brenner (1976).

[17]Contributions include Kriedte (1979), Berend (1986), Robisheaux (1998), Bideleux and Jeffries (2007), Süchs (1988), Maddalena (1977), Brenner (1976), Anderson (1974a), and Blum (1957).

[18]See Stoianovich (1994), Todorov (1983), Sugar (1977), Bairoch (1988), Hohenberg and Lees (1985), de Vries (1984), Bideleux and Jeffries (2007), and Anderson (1974b).

4 How Zipf's Law Emerged

4.1 Documenting the Facts

In this section I use graphs, OLS, quantile, and robust regression to document the paper's central motivating fact: that Zipf's Law emerged over time.

Figure 2 provides the motivating picture for this paper. It describes 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).[19] Table 2 measures the historical deviations from Zipf's Law. It provides quantitative evidence that deviations from Zipf's Law went from being large in 1300 to small in 1800.

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 (2008) 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.[20] 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 presents parameter estimates from (4). It shows that we can reject Zipf's Law in Western Europe up through 1500, but that we can not reject Zipf's Law in Western Europe from 1600 forwards.

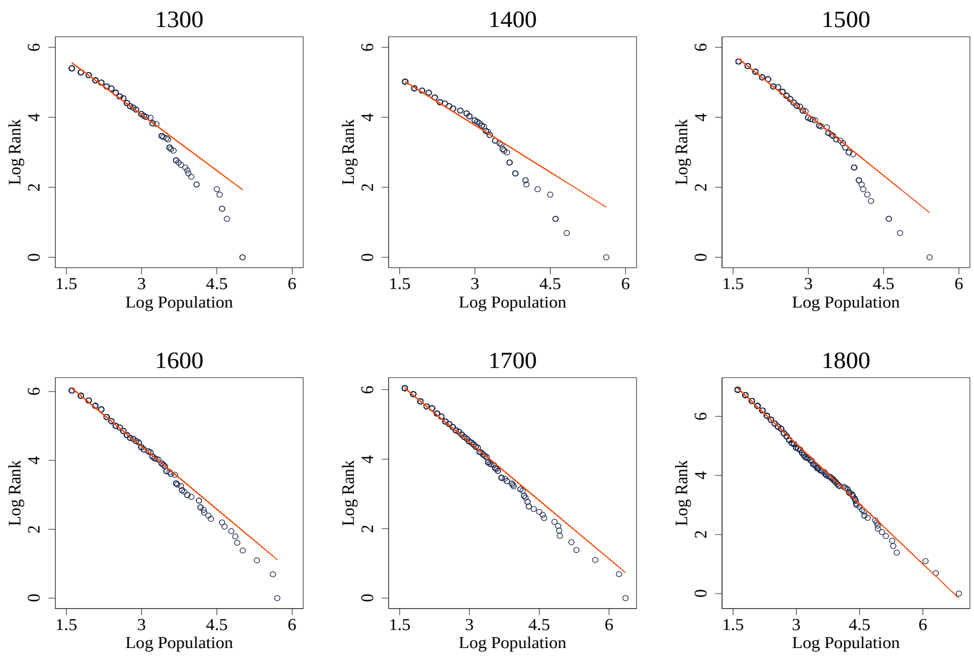
Quantile regression identifies more precisely where over the range of city sizes the curvature in the rank-size relation emerges.[21] Table 4 presents historical estimates of local,

[19]The way robust regression can be used to gauge departures from power laws is discussed below. Appendix D discusses the Theil estimator and shows that for estimating power law exponents it is superior to OLS and competitive with the adjusted-OLS estimator proposed by Gabaix and Ibragimov (2011) in terms of both small sample properties and precision.

[20]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 (2008), 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 E presents results from this specification which support the conclusion that Zipf's Law emerged in Western Europe 1500-1800.

[21]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

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



Note: 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 Zipf's Law

Year	Deviation
(1)	(2)
1300	6.27%
1400	4.21%
1500	1.51%
1600	0.58%
1700	0.50%
1800	0.18%

Note: 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: $MSD = N^{-1} \sum_{i=1}^{N} (S_i^a / S_i^z - 1)^2$.

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 4, as τ declines the estimates

estimates assume a piecewise linear loss function and minimizes the (asymmetric except in the case where $\tau = 0.5$) sum of absolute residuals. See Koenker (2005).

describe the local Zipf exponents (slopes) associated with progressively larger cities.[22] 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 from 0.9 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 (i.e. as τ declines) and fall within the (0.7, 1.5) range observed in contemporary economies (Soo 2005).

Table 3: A Regression-Based Test of Deviations from Zipf's Law

Year	Observations	δ_1	δ_2	Reject ZL
(1)	(2)	(3)	(4)	(5)
1300	255	-1.31	-0.33	Yes
		(0.02)	(0.03)	
1400	187	-1.15	-0.26	Yes
		(0.12)	(0.03)	
1500	321	-1.35	-0.24	Yes
		(0.11)	(0.02)	
1600	514	-1.33	-0.11	
		(0.08)	(0.01)	
1700	539	-1.22	-0.06	
		(0.07)	(0.00)	
1800	1,311	-1.40	-0.02	
		(0.05)	(0.00)	

Note: The regression 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.

4.2 Ruling Out Measurement Error as an Explanation

The observed deviations from Zipf's Law cannot be plausibly accounted for by non-classical measurement error or missing data.[23] I document this by showing that the population shortfalls in the upper tail are so big that they cannot be due to undercounting. I also show that if missing or mismeasured data for small cities were to account for observed deviations from

[22] 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.

[23] As noted above, classical measurement error will not account for the observed deviations from Zipf's Law. As shown in Appendix B, a comparison of Bairoch data to the data in de Vries (1986) reveals no evidence that big city populations are systematically mismeasured in the Bairoch data. In addition, the observed deviations from Zipf's Law are not explained by the fact that the Bairoch data round populations to the nearest thousand.

Table 4: Quantile Regression Estimates of Zipf Exponents

Year	Quantile Slope Parameters				
	τ = 0.9	τ = 0.75	τ = 0.5	τ = 0.25	τ = 0.1
(1)	(2)	(3)	(4)	(5)	(6)
1500	1.12	1.17	1.19	1.16	1.42
	(0.04)	(0.01)	(0.01)	(0.04)	(0.09)
1600	1.26	1.28	1.24	1.25	1.32
	(0.01)	(0.01)	(0.02)	(0.04)	(0.01)
1700	1.12	1.12	1.13	1.19	1.24
	(0.01)	(0.00)	(0.02)	(0.01)	(0.01)
1800	1.33	1.37	1.39	1.39	1.41
	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)

Note: Quantile slope parameter $\beta(\tau)$ 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. Bootstrapped standard errors in parentheses.

Zipf's Law this would imply implausibly high urbanization rates 1300-1500. In section 5, I show that the observed deviations from Zipf's Law are consistent with a rich body of historical evidence on the demography of European cities.

To gauge the possibility that the data undercount populations in the largest cities, I estimate hypothetical Zipf's Laws and calculate deviations from these benchmarks. The exercise amounts to asking: 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 (Appendix D provides documentation). 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. Further, examination of the residuals from a robust regression can identify outliers (Koenker 2005).

Tables 5 and 6 show that the magnitudes of the city population shortfalls in the upper tail are so large that non-classical measurement error is not a plausible explanation for the observed deviations from Zipf's Law. Table 5 uses the Theil estimator to construct a measure of deviations from Zipf's Law. It uses the Theil regression predictions displayed in Figure 2 to compare observed population to "Zipf-consistent" population for the biggest cities in Western Europe.[24] It shows that between 1500 and 1700 the biggest cities were far smaller than they

[24]This measure of deviations is conservative. As noted above, the data record populations in thousands. As a result, a number of cities that almost certainly had different populations are recorded as having the same number of inhabitants. The data record that Troyes, Siena, Strasbourg, and Louvain all had 17,000

needed to be to satisfy a rank-size rule. The fact that the divergences are all shortfalls is consistent with the narrative evidence presented in the next section. Table 6 shows actual and Zipf-consistent populations for the top 10 cities in 1500. It documents that Paris and Naples needed to be three times larger than they were in 1500 – and, counterfactually, as approximately as populous as they were in 1800 – to conform to Zipf's Law.

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 in periods where we observe deviations from Zipf's Law. For instance, if we believe that the populations of the top 10 cities are correctly measured and that a power law holds, we can estimate the implied populations of all subsequent cities in the urban hierarchy. This exercise implies European urbanization rates in 1300 that – implausibly – equal observed urbanization rates in 1700 (i.e. 14 percent of the population in cities with 5,000 inhabitants).

Table 5: The Ratio of Actual to Zipf-Consistent Population

Top 10 Cities	1500	1600	1700	1800
(1)	(2)	(3)	(4)	(5)
1	0.3	0.4	0.5	1.1
2	0.3	0.7	0.8	1.1
3	0.4	0.7	0.7	1.2
4	0.4	0.6	0.6	0.7
5	0.4	0.7	0.7	0.8
6	0.4	0.8	0.6	0.9
7	0.5	0.8	0.7	0.8
8	0.5	0.8	0.8	0.8
9	0.5	0.8	0.8	0.8
10	0.5	0.7	0.7	0.9

Note: This table shows the ratio of actual population (S_i^a) to Zipf-consistent population (S_i^z). Zipf-consistent population is estimated using the predicted values from the Theil regressions in Figure 2.

inhabitants, and by implication a common size rank of 81, in 1400. One way to address this problem is to break ties in ranks – say, by adding arbitrary small noise to all populations before ranking. The substantive results and conclusions do not change when one does this.

Table 6: Zipf-Consistent Population in 1500 and Observed Population in 1800

City	Zipf-Consistent Population 1500	Observed Population 1800
(1)	(2)	(3)
Paris	690	550
Naples	378	430
Venice	266	138
Milan	266	135
Granada	171	70
Lisbon	146	195
Tours	128	13
Genoa	114	90
Palermo	103	139
Gent	103	55
Total	2,365	1,815

Note: Population figures for the ten most populous cities as of 1500, in thousands. Zipf-consistent population is estimated using the predicted values from the Theil regressions in Figure 2.

4.3 Implications

The observed deviations from Zipf's Law have three key implications. First, Zipf's Law is due to something beyond simply a power law distribution of propitious locations. This conclusion follows from the fact that Zipf's Law emerged between 1500 and 1800, while the principal geographic features of the European landscape – e.g the location of navigable rivers and bays suitable for ports – remained essentially unaltered before 1800.[25]

Second, 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.

Third, something other than specialization in goods production accounts for Zipf's Law. Models of urban hierarchies, from Henderson (1974) to Black and Henderson (1999), Cordoba (2004), and Rossi-Hansberg and Wright (2007), assume industrial specialization accounts for city size distributions. 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 for their particular activities. But Figure 2 shows that "modern" patterns of 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 was relatively limited.[26]

[25]That Zipf's Law emerged over time is similarly problematic for models with an equilibrium distribution of populations due to fixed distributions of talent, agglomeration economies, and congestion costs.

[26]Nicholas (2003: 7) observes that, "Probably no pre-modern city was as functionally specialised as modern

5 Explaining the Emergence of Zipf's Law: History

In this section I discuss why 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. I also discuss the demography of pre- and early modern cities.

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 (Pounds 1990; Braudel 1979a; Bairoch 1988). Contemporaries recognized that this constraint prevented the proportionate growth (size-independent growth rates) associated with Zipf's Law. In 1602, Giovanni Botero noted that, "cities once grown to a greatness increase not onward according to that proportion." Botero considered and rejected explanations centered on wars, plagues, and chance. He observed that the absence of proportionate growth was explained by the difficulty large cities had in feeding themselves given prevailing transport costs (Botero 1602, Book 2, Pt. 9).

In early modern Europe, secure access to food supplies was a precondition of the growth of large cities: agricultural surpluses were limited and poor harvests brought famine. The security of supplies was often dependent on the degree of control cities could exert on the surrounding countryside (Scott 2004; Braudel 1979a). For Paris, the largest city in 17th century Europe, the problem of securing foodstuffs was especially acute, and is repeatedly stressed by contemporary commentators (Pounds 1990). In 1591, Pope Gregory XIV issued an edict designed to facilitate the provisioning of Rome from its countryside. In Northern Italy, great cities – like Milan and Florence – conquered and dominated dependent territories that included smaller cities and agricultural hinterlands (Chittolini 1994). Cities on the Istrian and Dalmatian coast similarly controlled territories that stretched inland to the mountains (Vilfran 1994). 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 (Scott 2004). 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.[27]

industrial cities tend to be."

[27]See Scott (2005), de Vries (1976), Pounds (1990), Weber (1958), Pirenne (1958), Chittolini (1994), Livet (2003), Blockmans (1994), Nicholas (2003), Vilfran (1994), Miller (2008), Braudel (1979c).

In addition to control over the countryside, transportation costs were a central constraint. Transportation costs – especially for heavier products and overland transport – were exceedingly high. 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.[28] As a result, cities preserved substantial forms of land-intensive production. There were gardens, fields, and areas devoted to livestock within cities themselves.[29] Costs associated with the transport of fuel generated similar bottlenecks (Ballaux and Blondé 2004).

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, respectively, 89 and 270 percent higher in Antwerp (commercial hub of the relatively urbanized Low Countries) than in Danzig (principal port of rural Poland). Pounds (1979: 61) observes that prices of agricultural products were increasing in town size. The data support this observation and the argument that, while food prices were increasing in city size, the land constraint softened over time. Figure 6 plots consumer prices and bread prices from Allen (2001) against city population, along with the fitted values from a median regression of consumer prices on city size.[30] It shows that (1) consumer prices tracked bread prices, (2) prices were associated with city size, and (3) that the price gradient declined over time.[31]

Three factors relaxed the land constraint on city growth: the development of the international grain trade, improvements in agricultural productivity, and increases in productivity in mercantile and knowledge-based activities that concentrated in larger cities.

The Baltic grain trade was known to the Dutch as the *moedernegotie*: the "mother of all trades" (Bogucka 1980; Allen and Unger 1990; van Tielhof 2002). The emergence of the Baltic trade in grain can be clearly dated to the mid-15th century (Davies 1982: 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

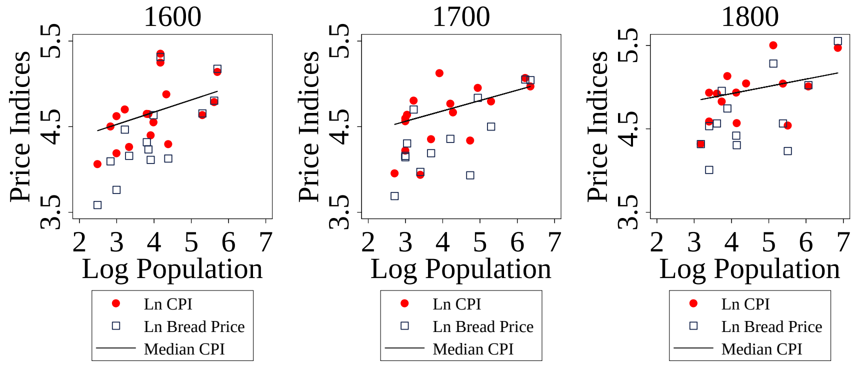
[28] See Pounds (1979: 61), Nicholas (2003: 43), and Braudel (1979a: 133). Ballaux and Blondé (2004) suggest transport over land was four times more expensive than on navigable rivers.

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

[30] Basic land-intensive products account for 2/3's of spending in Allen's (2001) consumer price index.

[31] The OLS estimate of the relationship between log prices and log city populations (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.

Figure 6: Consumer Prices and City Sizes



Note: Price data are from Allen (2001). Bread prices scaled by multiplying by 100.

on grain imports from the Baltic region. 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). As shown in Table 7, Dutch grain imports were sufficient to feed the entire urban population of the Netherlands by 1600. Moreover, as de Vries and van der Woude (1997: 414-5) observe, "grain was the commodity that gave Dutch merchants entrée to the Iberian and Mediterranean ports from the 1590s on." In the late 1500s, Iberian and Italian cities facing local resource constraints were dependent on imports of Baltic grain (Allen and Unger 1990; Bogucka 1978).

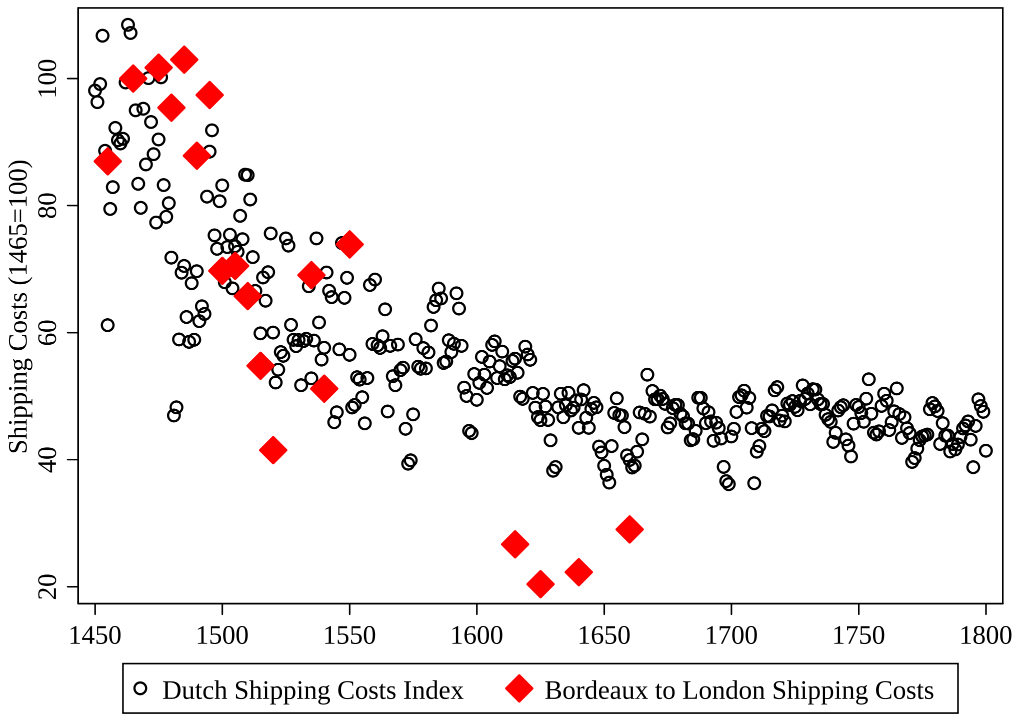
Table 7: Dutch Imports of Baltic Grain

Period	Hectolitres/Year	Enough to Feed
(1)	(2)	(3)
1550s	1,053,500	263,375
1590s	1,745,800	436,450
1640s	2,859,500	714,875

Note: Import data are from van Tielhof (2002). Calculations assume consumption of 4 hectolitres per person per year. In 1600, the 24 largest Dutch cities had 421,000 inhabitants.

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; Unger 2007; 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). Figure 7 documents significant declines in the cost of Dutch shipping, and similar reductions in the cost of shipping wine from Bordeaux to London through the 1600s. 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 Zanden and van Tielhof 2009). However, similar increases in productivity and declines in costs are observed in trans-Atlantic shipping. Menard (1991) finds productivity growth of 1.4% per year in shipping rice from Charleston, South Carolina to London 1700-1776 and almost identical rates of productivity growth in trans-Atlantic tobacco shipping over even longer periods.[32]

Figure 7: Real Maritime Transport Costs



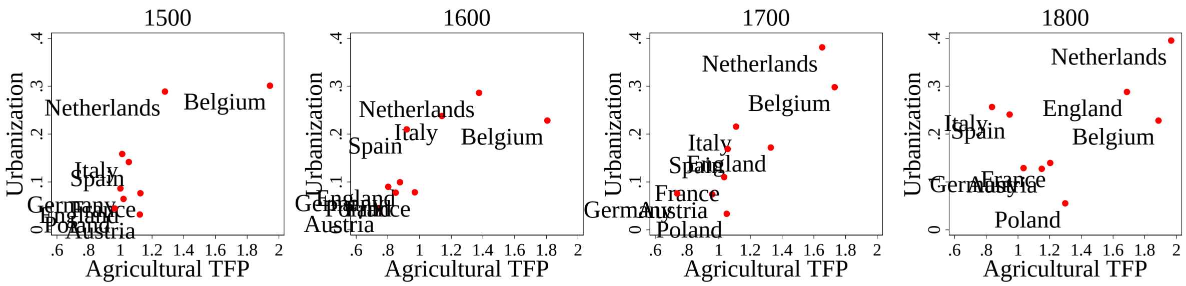
Note: The Dutch shipping cost index is from van Tielhof and van Zanden (2009). Nominal Dutch shipping costs are deflated by the wholesale price index also from van Tielhof and van Zanden (2009). Bordeaux-London shipping costs are from Menard (1991) and describe the costs associated with transporting tons of wine from France to England. Real Bordeaux-London shipping costs are obtained by deflating the nominal cost series by the English consumer price index also from Menard (1991).

Increases in agricultural productivity also relaxed the land constraint, and were a key factor in the growth of cities in Northwest Europe (Maddalena 1977, Pounds 1990, Kriedte 1979). Agricultural productivity was positively associated with urbanization, and economies where city growth was concentrated experienced relatively high productivity growth in agri-

[32] Data for periods before 1450 are limited, in part because international transactions that emerged 1450-1600 were often 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.

culture. Figure 8 shows how urbanization rates were correlated with estimates of agricultural total factor productivity in nine macroeconomies.

Figure 8: Agricultural TFP and Urbanization in European History



Note: Agricultural TFP estimates from Allen (2003). Urbanization rates from Acemoglu et al. (2005).

Over the period 1500-1700 increases in productivity in mercantile and knowledge-based activities concentrated in larger cities also relaxed the land constraint. Atlantic ports benefitted from new trade routes connecting them to both the Atlantic and the East Indies (Acemoglu et al. 2005). The information revolution associated with the diffusion of the printing press raised productivity in business, administration, and educational activities concentrated in larger cities. Printing presses were typically established in larger cities and promoted activities that were intensive in human capital as opposed to land. The establishment of printing presses was associated with significant increases in city growth. Cities with printing presses grew as much as 60% faster than otherwise similar cities 1500-1600 (Dittmar 2011). These growth effects were concentrated at the upper end of the city size distribution.

While the period 1500-1800 saw increases in agricultural productivity, the trade in food, and the scale of knowledge-intensive activities, the public health environment in cities did not improve substantially. An extensive literature on the demography of early modern cities finds that urban death rates exceeded urban birth rates and rural death rates.[33] In general, mortality increased in city size (Woods 2003). 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 well into the 1700s. It was only in the wake of the work by John Snow in the mid-1850s that germ theories of disease transmission began to be accepted. Previously, the dominant view was the miasma theory – which held that diseases such as cholera or the Black Death were caused by a form of "bad air." These facts support the conclusion that the key changes were not in the domain of public health.

[33]See Woods (2003), Mols (1955, 1956), de Vries (1984), Bairoch (1988), and McIntosh (2001).

6 Explaining the Emergence of Zipf's Law: Model

6.1 Motivation

The 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.[34] 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).

6.2 Environment

The model has overlapping generations. At any time t, cities indexed with i have old residents N^o_{it} and arriving young residents N^y_{it}, 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.[35]

There are city-specific shocks to either amenities or productivity. Without loss of generality, assume cities are subject to amenity shocks a_{it} due to some combination of policy and nature. In particular:

$$a_{it} = \epsilon_{it}(1 - \tau_{it}) \tag{5}$$

ϵ_{it} is an iid city-specific shock and $\tau_{it} \in (0, 1)$ is a city-specific distortion. Without loss of

[34] 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.

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

generality, the amenity shocks a_{it} enter utility multiplicatively:

$$u(c) = a_{it}c \tag{6}$$

Production is Cobb-Douglas in technology (A), labor (young N^y and old N^o), and land (L)[36]:

$$Y_{it} = A_{it}(N^y_{it})^\alpha (N^o_{it})^\beta (L_{it})^{1-\alpha-\beta} \tag{7}$$

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.[37] The wage is the marginal product of labor and is consumed in each period:

$$c^j_{it} = w^j_{it} = \frac{\partial Y_{it}}{\partial N^j_{it}} \quad j \in \{y, o\} \tag{8}$$

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.[38]

6.3 Analysis of City Growth – The General Case

Young people choose a city i subject to city-specific taxes τ_{it} and given the existing distribution of populations (wages). Their individual maximization problem reduces to[39]:

$$\max_i \, a_{it}w^y_{it}$$

[36] 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).

[37] 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.

[38] 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.

[39] 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: $u_{it}^{y} = u_t$. It follows that:

$$w_{it}^{y} = \frac{u_t}{a_{it}} \tag{9}$$

Because young people earn wages equal to their marginal product, we have that:

$$w_{it}^{y} = \alpha A_{it} (N_{it}^{y})^{\alpha-1} (N_{it}^{o})^{\beta} (L_{it})^{1-\alpha-\beta} \tag{10}$$

Combining (5), (9), and (10), we get an expression for the number of new-comers in the representative city:

$$N_{it}^{y} = (N_{it}^{o})^{\frac{\beta}{1-\alpha}} (A_{it})^{\frac{1}{1-\alpha}} (L_{it})^{\frac{1-\alpha-\beta}{1-\alpha}} (1-\tau_{it})^{\frac{1}{1-\alpha}} \left(\frac{\alpha \epsilon_{it}}{u_t}\right)^{\frac{1}{1-\alpha}} \tag{11}$$

The representative city growth rate is:

$$g_{it}^{N} \equiv \frac{\Delta N_{it}}{N_{it}} = \frac{N_{it}^{y} - \delta N_{it}^{o}}{N_{it}^{o}} \tag{12}$$

Substituting with equation (11) gives:

$$g_{it}^{N} = (N_{it}^{o})^{\frac{\beta+\alpha-1}{1-\alpha}} (A_{it})^{\frac{1}{1-\alpha}} (L_{it})^{\frac{1-\alpha-\beta}{1-\alpha}} (1-\tau_{it})^{\frac{1}{1-\alpha}} \left(\frac{\alpha \epsilon_{it}}{u_t}\right)^{\frac{1}{1-\alpha}} - \delta \tag{13}$$

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

6.4 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 (13) gives:

$$g_{it}^{N} = (A_t)^{\frac{1}{1-\alpha}} \left(\frac{\alpha \epsilon_{it}}{u_t}\right)^{\frac{1}{1-\alpha}} - \delta \tag{14}$$

Since the only city-specific argument on the right-hand side of (14) 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).[40]

6.5 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 to begin assume that $A_{it} = A_t$. Assume also no distortions: $\tau_{it} = 0$. We now have the following variant of equation (13):

$$g_{it}^N = (N_{it}^o)^{\frac{\beta+\alpha-1}{1-\alpha}} (A_t)^{\frac{1}{1-\alpha}} \left(\frac{\alpha\epsilon_{it}}{u_t}\right)^{\frac{1}{1-\alpha}} - \delta \tag{15}$$

Here land has a positive income share because $\alpha + \beta < 1$. This fact secures the key feature of equation (15): 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.

The logic of city growth with a fixed factor is illustrated with a simple simulation. For simplicity, I take the income shares for young and old workers to be $\alpha = 0.3$ and $\beta = 0.6$, implying that land's share in production is 0.1.[41] I let all cities start with the same fixed amount of land, normalized so that $L_{it} = 1$ for all cities and time periods. Finally, I select an arbitrary initial distribution of city populations.

A simple simulation using the model generates two principal findings. First, city populations settle into a non-Zipf distribution broadly similar to the city size distribution observed in Europe in 1300. Second, there is non-random growth.

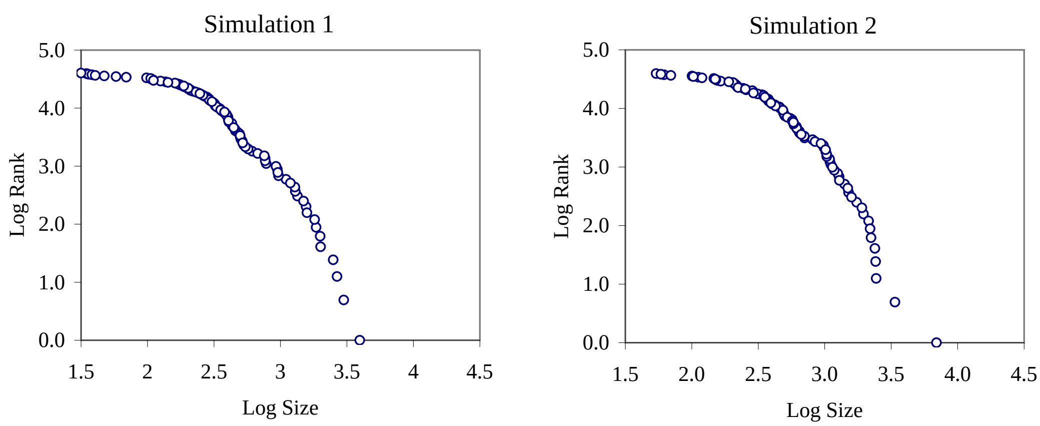
Figure 8 shows the city-size distributions that result when one takes fixed set of cities of arbitrary starting size and runs them through the model assuming that the fixed land

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

[41] The broad results of the simulation are not sensitive to reasonable perturbations of these parameters.

(L) has a positive income share and that productivity is static and common across cities. The simulation is run over 250 periods. It is assumed that $\alpha = 0.6$, $\beta = 0.2$, $\delta = 0.1$. The scaling factor u is chosen to lend plausible final sizes, but has no impact on the shape of the distribution. With no technological change, the model tends to a state with no growth in population (or per capita income) aside from ephemeral variations induced by stochastic shocks. Simulating the model with taxes $\tau_{it} > 0$ and increasing in city size gives equivalent results. Over 250 periods ("years") the correlation between city growth and city size is approximately -0.2. As documented in the next section, this magnitude is consistent with the the quantitative historical data from periods before the land constraint was relaxed.

Figure 8: Simulation of Model



Note: This figure presents simulated city populations and city rankings from two representative runs of the model over 250 periods.

The effects of the land constraint on city growth and on the distribution of city populations can be undone by falling transport costs or increases in agricultural productivity.[42] Similarly, that innovations that increase productivity in large cities – such as the opening of new trade routes that increase the productivity of merchants or the emergence of a technology like printing that historically had special applications in large cities – will also effectively relax the land constraint.

[42]For simplicity this paper presents a case with a land constraint and a case without. One could extend the model by positing that the land share in production is ϕ, that $\alpha + \beta + \phi = 1$, and that over time $\phi \to 0$.

7 Explaining the Emergence of Zipf's Law: Empirics

This section examines random growth and locational fundamentals theories of Zipf's Law.

7.1 Random Growth

Leading theories that account for Zipf's Law posit random growth. This section establishes when and how random growth emerged in Western Europe.

Table 8 shows that random growth emerged only after 1500. Before 1500, there was a significant negative correlation between size and subsequent growth in every period except 1300-1400, the century of the Black Death.[43] From 1500 forwards there is no significant correlation between size and growth.

Table 8: Correlations Between City Size and City Growth

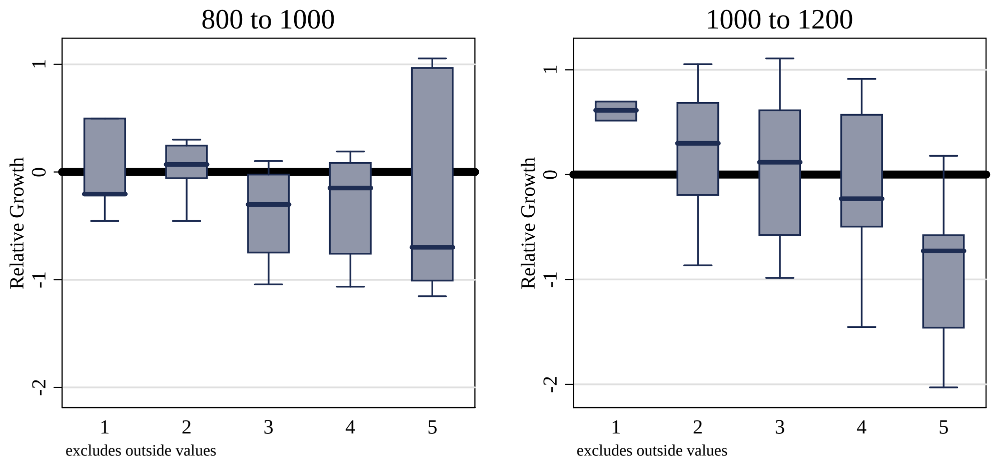
Period	Correlation
(1)	(2)
1000 to 1200	-0.69 **
1200 to 1300	-0.23 **
1300 to 1400	-0.09
1400 to 1500	-0.27 **
1500 to 1600	-0.05
1600 to 1700	0.00
1700 to 1800	-0.07

Note: This table presents correlations between normalized city sizes and growth rates. 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 $\hat{g}_{it} = (g_{it} - \bar{g}_t)/\sigma_t$. Significance at the 95 and 90 percent confidence levels denoted "**" and "*", respectively.

Next I group cities into size quantiles and examine the distribution of growth rates within each quantiles. Figure 9 presents box-plots of city growth by size quintile (quintile 1 comprises the smallest cities and quintile 5 the largest). It confirms that large cities were at a pronounced growth disadvantage 800-1200. Figure 10 presents data for 1200-1800 and shows how random growth emerged from 1500 forward. Figure 11 compares the distribution of growth rates for the top 10 percent and bottom 90 percent of cities using nonparametric

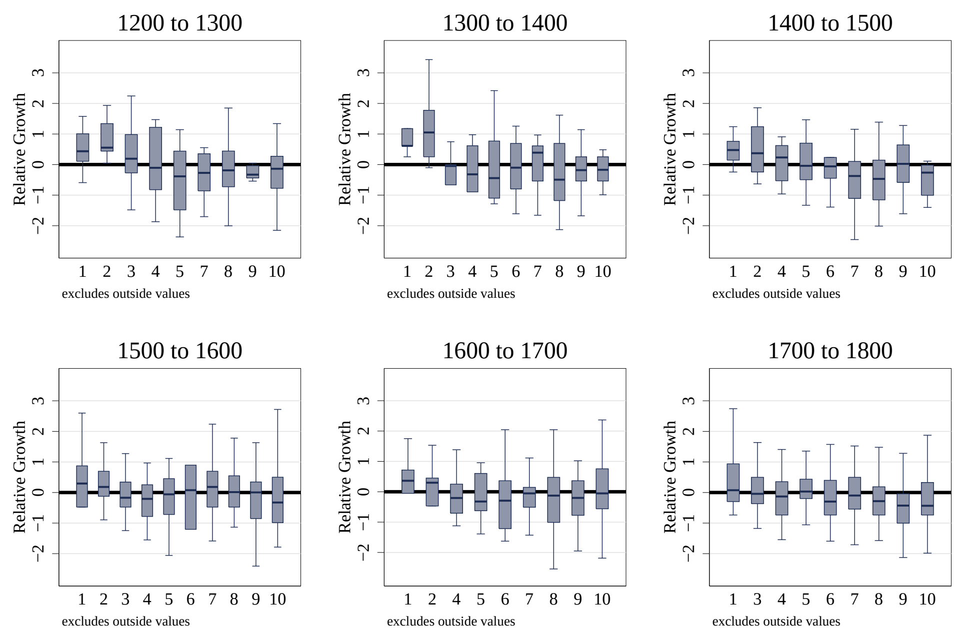
[43] The Plague epidemics that swept Europe in the mid-1300s killed approximately 1 in 3 people. As shown in Table 1 (above) city populations fell dramatically 1300-1400. The observed annualized growth rate implies a 20% decline in city populations 1300-1400.

Figure 9: City Growth Rates By Population Quintile 800-1200



Note: This figure documents how normalized growth rates varied with city size. The smallest cities are in quintile 1, the largest in quintile 5. The boxes describe the interquartile range. The line within each box is the decile's median growth rate. The "whiskers" mark the adjacent values. See note to Table 8 for calculation of normalized growth rates.

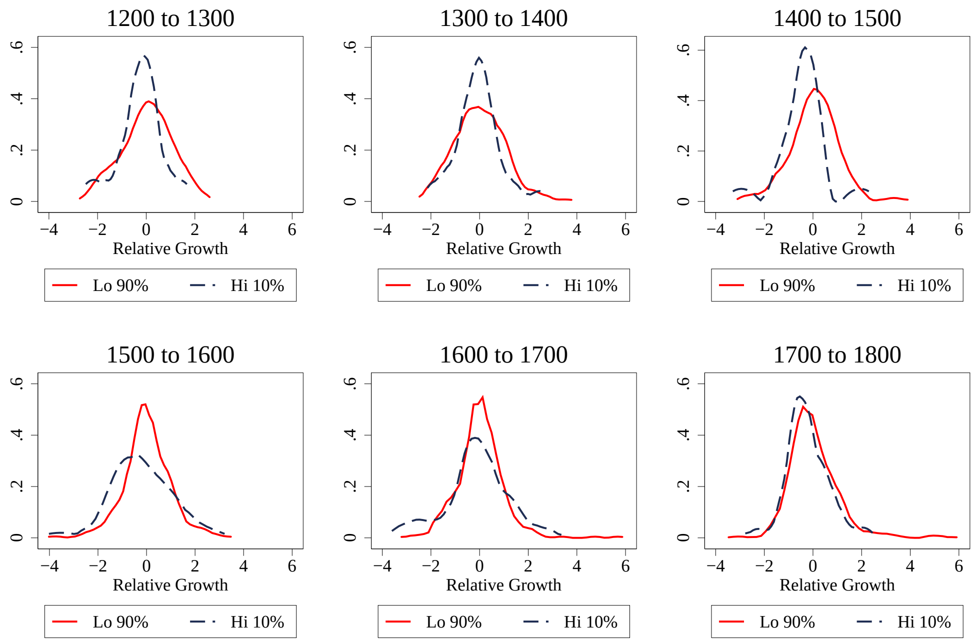
Figure 10: City Growth Rates By Population Decile 1200-1800



Note: The smallest cities are in decile 1, the largest in decile 10. The boxes describe the interquartile range. The line within each box is the decile's median. See note to Table 8 for calculation of normalized growth rates.

kernel densities. Figure 11 shows that up through the period 1500-1600, the largest cities consistently grew more slowly than smaller cities, and that by 1700-1800 large and small cities were drawing growth rates from approximately identical distributions.

Figure 11: The Distribution of City Growth Rates



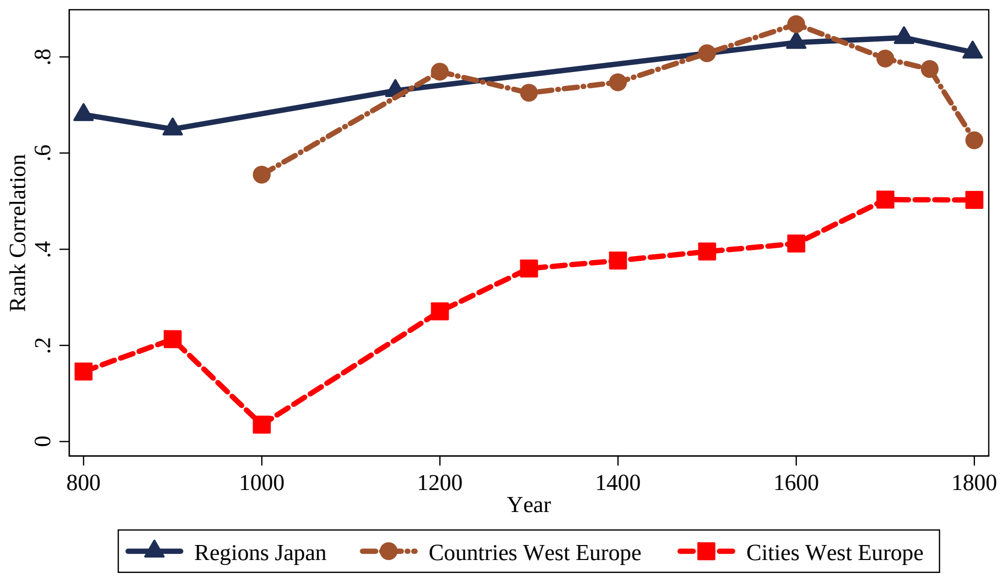
Note: This figure presents kernel densities of the distribution of growth rates for the largest 10 percent (Hi 10%)and smallest 90 percent (Lo 90%) of cities. See note to Table 8 for calculation of normalized growth rates.

7.2 Locational Fundamentals

Geographic theories of Zipf's Law hold that the distribution of natural advantages across locations is the underlying determinant of city population distributions. This paper documents that European city populations have not always obeyed a power law distribution. If the distribution of city populations is determined by locational fundamentals, this finding suggests that the fundamentals may be dynamic. For instance, the advantages of port locations may be magnified in an era of relatively cheap ocean-going transport or when new trade routes are opened. Locational advantages may also be quite literally constructed (harbors may be dredged, canals dug, etc.).

Two key findings have been used to support a geographic theory of Zipf's Law. The first is the observation that a regional analogue of Zipf's Law holds over many centuries in the Japanese data and that Japanese regional population densities in the past are highly correlated with contemporary population densities (Davis and Weinstein 2002). As Davis and Weinstein (2002) note, the observed high level of persistence raises the question of whether Japan was special. Figure 11 shows that the high correlations found in the Japanese regional data are not matched in the European data on city populations, but are matched in the European data on national population densities. This suggests that Japan was not special, but that a distinction between regional and city-level population data may be useful and that discussions of Zipf's Law should (data permitting) focus on cities.

Figure 11: Correlation Between Historic and Contemporary Populations



Note: Contemporary and historic city population data are from Brinkhoff (2008) and Bairoch, Batou, and Chèvre (1988), respectively. Contemporary national population data are for the year 2000 from Eurostat (2009) and Statistics UK (2001). Historic national population data are from Acemoglu, Johnson, and Robinson (2005). Rank correlations for regional and national data are calculated using population densities (i.e population divided by land area in square kilometers). Rank correlations for Japanese regions are from Davis and Weinstein (2002).

The second key finding for geographic theories is that city size hierarchies have been stable in the face of large temporary shocks in the 20th century. Davis and Weinstein (2002) and Brakman et al. (2004) document that extensive and selective bombing of German and Japanese city cities during the second war had little long run impact on the distribution of city populations in these countries.

Evidence on such quasi-natural experiments in the more distant past is very fragmentary:

we do not have high frequency data for European city populations around shocks such as the Black Death (1348-1350) or the Thirty Years War (1618-1648). However, the long run evidence suggests that historically European city size hierarchies were relatively dynamic.

The historical data reveal significant churning in European city hierarchies and support the conclusion that geography was not destiny in any direct sense. 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.[44]

8 Conclusion

Zipf's Law is supposedly one of the most robust empirical regularities in economics. This paper has shown that, to the contrary, Zipf's Law emerged over time in European history. In particular, Zipf's Law emerged over the transition to modern economic growth as city production became less reliant on quasi-fixed local land endowments. With developments in trade, agricultural productivity, and knowledge-intensive activities, city growth rates became random, in the sense of being independent of city population. This transformed the urban structure of Europe.

[44]For historical populations see Bairoch et al. (1988), Meigs (1973), and Stillwell et al. (1976). Contemporary French populations are for urban agglomerations and are from Brinkhoff (2008).

The historical emergence of Zipf's Law also 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 Zipf's Law suggests that locational advantages may emerge with economic development, and hence be endogenous along important dimensions. In addition, the fact that Zipf's Law emerged in an era when the industrial specialization of urban activity was relatively limited 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. The historical evidence is consistent with theories emphasizing random growth in the emergence of Zipf's Law.

References

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

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

[3] 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.

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

[5] Allen, Robert C. (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.

[6] Allen, Robert C. and Richard W. Unger (1990), "The Depth and Breadth of the Market for Polish Grain 1500-1800," in J.Ph.S. Lemmink and J.S.A.M. Koningsbrugge eds., *Baltic Affairs: Relations between the Netherlands and North-Eastern Europe 1500-1800*, Nijmegen; Instituut voor Noord- en Oosteuropeses Studies.

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

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

[9] 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.

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

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

[12] 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.

[13] Behrens, Kristian, Gilles Duranton, and Frédéric Robert-Nicoud (2010), “Productive cities: Sorting, selection and agglomeration,” working paper. On-line: http://individual.utoronto.ca/gilles/Papers/Sorting.pdf

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

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

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

[17] 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.

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

[19] 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.

[20] Bogucka, Maria (1980), “The Role of the Baltic trade in European Development from the 16th to the 18th Centuries,” *Journal of European Economic History*, Vol. 9, No. 1, pp. 5-20.

[21] Bogucka, Maria (1978), "North European Commerce as a Solution to Resource Shortage in teh Sixteenth-Eighteenth Cenuries," in A. Maczak and W.N. Parker eds., *Natural Resoures in European History*, Washington.

[22] 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.

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

[24] 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.

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

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

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

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

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

[30] 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 , AD 1000-1800*, Westview; Boulder.

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

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

[33] Cordoba, Juan (2004), "On the Distribution of City Sizes," Rice University Working Paper.

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

[35] 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.

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

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

[38] 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.

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

[40] 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.

[41] Dittmar, Jeremiah (2011), "Information Technology and Economic Change: The Impact of the Printing Press," forthcoming in *The Quarterly Journal of Economics*.

[42] Dittmar, Jeremiah (2010), "Cities, Institutions, and Growth: The Emergence of Zipf's Law," working paper. On-line at www.jeremiahdittmar.com/research.

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

[44] 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

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

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

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

[48] Gabaix, Xavier (2008), "Power Laws in Economics and Finance," NBER Working Paper 14299.

[49] 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.

[50] 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.

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

[52] Guérin-Pace, France (1995), "Rank-size distribution and the process of urban growth," *Urban Studies*, vol. 32, pp. 551 562..

[53] 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.

[54] Henderson, J.V. (2005), "Urbanization and Growth," in P. Aghion and S. Durlauf eds., *Handbook of Economic Growth*, Volume 1B, New York; Elsevier.

[55] Henderson, J.V. (2003), "The Urbanization Process and Economic Growth: The So-What Question," *Journal of Economic Growth*, Vol. 8, No. 1, pp. 1381-4338.

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

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

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

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

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

[61] Iaonnides, Yannis, Henry Overman, Esteban Rossi-Hansberg, and Kurt Schmidheiny (2007), "The Effect of Information and Communication Technologies on Urban Structure," CESIFO Working Paper No. 2049.

[62] Jacks, David (2004), "Market Integration in the North and Baltic Seas, 1500-1800," *Journal of European Economic History*, Vol. 33, No. 3, pp. 285-329.

[63] Jedin, Hubert, Kenneth Scott Latourette, Jochen Martin (1970), *Atlas Zur Kirchen-Geschichte*, Freiburg; Herder.

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

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

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

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

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

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

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

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

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

[73] 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.

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

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

[76] Menard, Russell (1991), "Transport Costs and Long-Range Trade 1500-1800: Was there a European 'Transport Revolution' in the Early Modern Era?" in James Tracy ed., *The Political Economy of Merchant Empires*, Cambridge; Cambridge University Press.

[77] 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.

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

[79] Özmucur, Süleyman and Sevket Pamuk, "Did European Commodity Prices Converge during 1500-1800?" Paper presented to the Conference "The New Comparative Economic History" in honor of Jeffrey G. Williamson. Online: http://www.econ.upenn.edu/ ozmucur/papers/ozmucur-pamuk-JeffFest-revised-%20May%202006.pdf

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

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

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

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

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

[85] 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.

[86] Rosenfeld, Hernán, Diego Rybski, Xavier Gabaix, and Hernán Makse, "The Area and Population of Cities: New Insights from a Different Perspective on Cities," *American Economic Review*, forthcoming.

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

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

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

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

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

[92] Siegal, A.F. (1982), "Robust regression using repeated medians," *Biometrika*, Vol. 69, No. 2, pp. 242-4.

[93] 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.

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

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

[96] 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 d'Italie (XIIIe-XVIe sièclea)*, Turnhout; Brepols Publishers.

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

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

[99] 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.

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

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

[102] 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.

[103] 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.

[104] 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.

[105] Vilfran (1994), in Charles Tilly and Willem Blockmans eds., *Cities and the Rise of States in Europe , AD 1000-1800*, Westview; Boulder.

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

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

[108] Wright, William (1966), *Serf, Seigneur, and Sovereign: Agrarian Reform in Eighteenth-Century Bohemia*, Minneapolis; University of Minnesota.

[109] 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.

— Not For Publication —

A Appendix: Data

Data on city populations: Historic city populations are from Bairoch, Batou, and Chèvre (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/.

Other data: National-level urbanization rates are from Acemoglu, Johnson, and Robinson (2005). Historical population densities are computed using national populations from Acemoglu, Johnson, and Robinson (2005). Contemporary population densities are computed using population data from Eurostat (2009) and Stastics UK (2001). Historical estimates of national-level agricultural TFP are from Allen (2003). Data on consumer prices in European cities are from Allen (2008).

B Appendix: Measurement Error

Given the deviations from Zipf's Law observed in the Bairoch data, it is natural to wonder whether the Bairoch data are consistent with alternate sources of historical data. This section compares the Bairoch data to the most comprehensive independent source for city population data, the database in de Vries (1984).[45] The Bairoch data covers all European cities that reached 5,000 inhabitants by or before 1800 and contains observations on 2,204 cities. The Bairoch data records populations every 100 years 800-1700 and every 50 years 1700-1850. 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 every 50 years 1500-1800.

Table B1 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

[45] The main body of the paper provides tests for measurement error in several additional ways. Section 4.1 shows that the data would have to embody implausibly large non-classical measurement error for Zipf's Law to have actually held. Section 5 documents that the observed deviations are consistent with the narrative evidence.

from 0.89 in 1500 to 1.00 in 1800; the ratio of recorded values approaches 1 and its standard deviation falls.

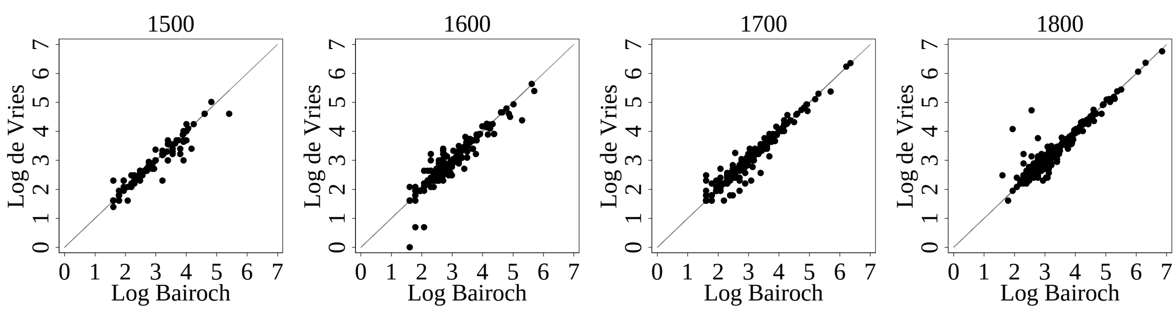
Table B1: Comparison of Source Data on City Populations

			Ratio of Bairoch Data to de Vries Data				
Year	Cities	Corr.	Mean	St. Dev.	Min.	Max.	Skew.
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
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

Note: This table compares population data from Bairoch et al. (1988) and de Vries (1984). Column (3) presents the correlation between recorded values. Columns (4) to (8) examine the ratio of these values.

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 B1 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.[46]

Figure B1: Comparison of Source Data on City Populations



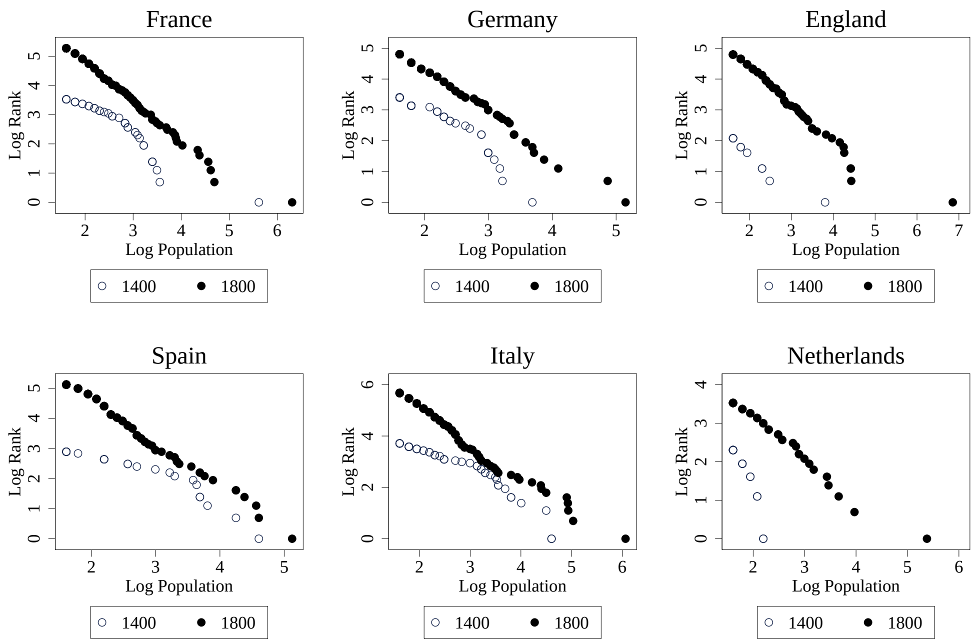
Note: This figure plots city populations recorded in de Vries (1984) against corresponding values in Bairoch et al. (1988). The 45 degree line is shown to clarify where the Bairoch data provide larger (smaller) values.

[46]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 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.

C Appendix: Emergence of Zipf's Law at the National Level

Data in Russell (1972) on urban systems in the high middle ages shows that Zipf's Law did not hold at the local level. Figure C1 shows how Zipf's Law emerged between 1400 and 1800 in the six leading national economies of Western Europe. Figure C2 shows how Zipf's Law

Figure C1: The Emergence of Zipf's Law at the National Level



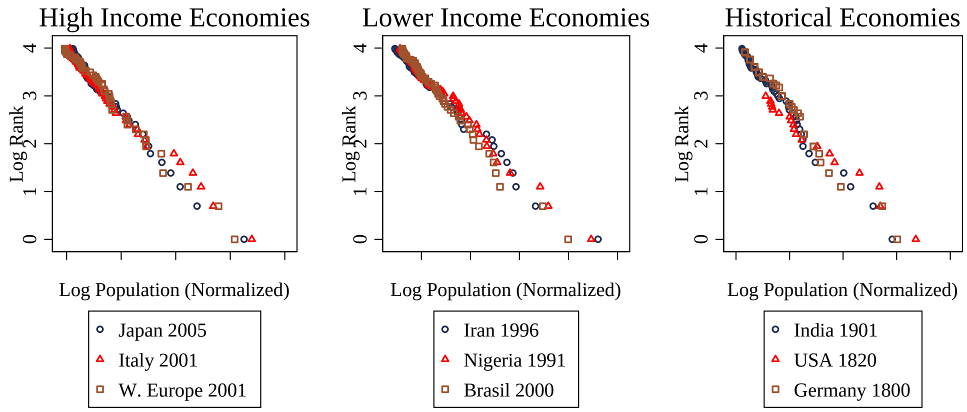
is apparent in data on city populations in three high income, three lower income, and three historical economies.

D Appendix: 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

Figure C2: Zipf's Law Across Time and Space



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

form:

$$\ln R_i = \alpha - \beta \ln S_i + \epsilon_i \tag{16}$$

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 (2007) 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.

$$\ln(R_i - 1/2) = \alpha - \beta \ln S_i + \epsilon_i \tag{17}$$

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

However, the second problem with least squares is that any OLS estimator may be subject to gross errors in contexts marked by significant outliers. This is because the OLS estimator suffers from sensitivity 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 here.

The literature has discussed the Hill maximum likelihood estimator (MLE) as an alternative to OLS.[47] However, as Gabaix and Ioannides (2004) observe, the small sample biases

[47]For a sample of n cities with sizes S_i ordered so that $S_{(1)} \geq \ldots \geq S_{(n)}$, the Hill estimator is: $\hat{\beta}_H =$

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 exploited in the Zipf's Law literature.[48] 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 $S_{ij} = (Y_j - Y_i)/(x_j - x_i), 1 \leq i < j \leq n$. 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 (2007) in eliminating small sample bias. This is evident in Figure D1, 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.[49] Figure D1 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 D1 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.

The Theil estimator also generates relatively precise estimates. Gabaix and Ibragimov (2007) show that, when we estimate power law exponents in small samples, OLS standard

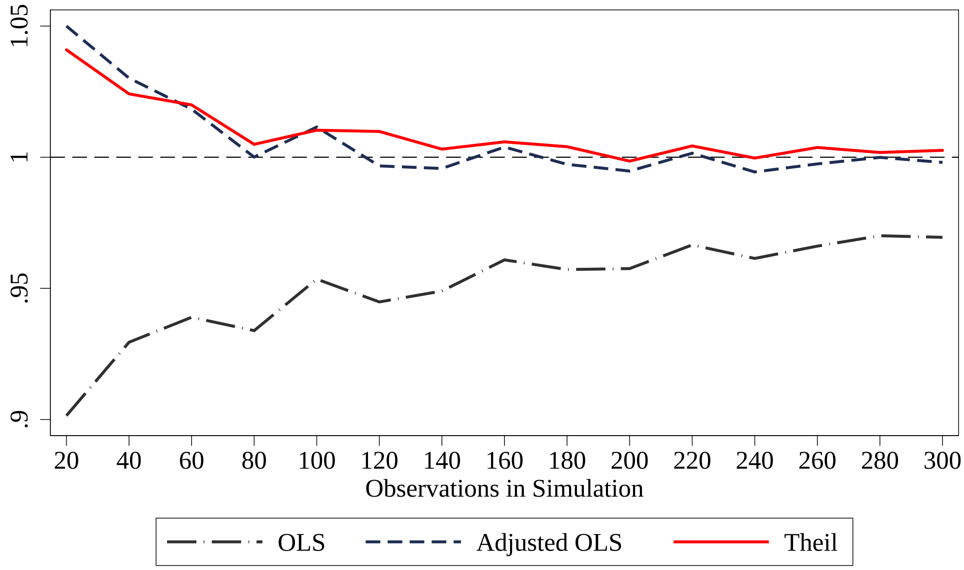
---

$(n-1)/\sum_{i=1}^{n}\left[\ln(S_{(i)}) - \ln(S_{(i+1)})\right]$.

[48]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.

[49]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.

Figure D1: Monte Carlo Estimates of the Zipf Exponent
Mean Estimates Over 1,000 Simulations



errors are biased down.[50] The confidence interval associated with Theil regression estimates similarly overstates the estimator's precision when data are drawn from a distributional power law.[51] To gauge and compare the true precision of these estimators, we can use Monte Carlo simulations. Figure D2 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.

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:

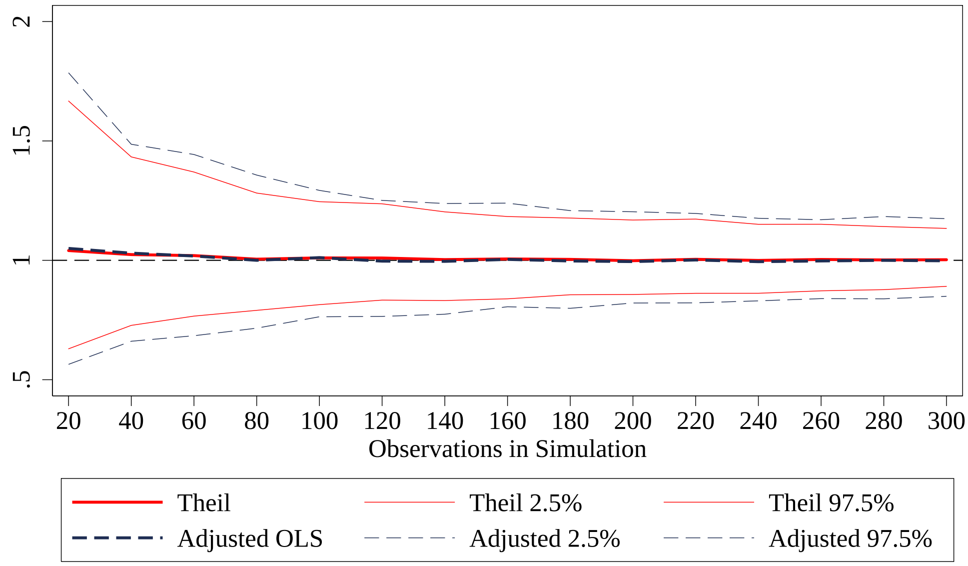
$$h \equiv (i,j) \qquad X(h) \equiv \begin{bmatrix} 1 & x_i \\ 1 & x_j \end{bmatrix} \qquad y(h) \equiv \begin{bmatrix} y_i \\ y_j \end{bmatrix} \qquad b(h) \equiv X(h)^{-1} y(h)$$

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."

[50] The true standard error of $\hat{\beta}$ in equation (17) is asymptotically $(2/n)^{0.5}\hat{\beta}$.

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

Figure D2: Monte Carlo Estimates of Zipf Exponent
Mean Estimate and 95% Interval Over 1,000 Simulations



E Appendix: Conventional OLS Regression Test

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:

$$\ln R_i = \alpha - \beta \ln S_i + \epsilon_i \tag{18}$$

A number of studies suggest employing a regression augmented with a quadratic term to detect non-linearities and deviations from distributional power laws[52]:

$$\ln R_i = \beta_0 - \beta_1 \ln S_i + \beta_2 (\ln S_i)^2 + \nu_i \tag{19}$$

As discussed below, the standard errors associated with this model are biased down. However, I present historical estimates of equation (19) to facilitate comparison with existing studies using non-historical data. Table 4 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.

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

Table E1: OLS Regression Analysis of Deviations from Zipf's Law

Year	Observations	Parameter β_1	Parameter β_2
(1)	(2)	(3)	(4)
1300	255	0.30	-0.28
		(0.08)	(0.02)
1400	187	-0.13	-0.22
		(0.20)	(0.04)
1500	321	0.20	-0.20
		(0.11)	(0.02)
1600	514	0.82	-0.08
		(0.04)	(0.01)
1700	539	0.95	-0.04
		(0.05)	(0.01)
1800	1,311	1.36	0.00
		(0.04)	(0.01)

Note: 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. Heteroskedasticity-robust standard errors in parentheses.

However, the estimates in Table E1 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 (19) exhibit downward bias in finite samples.[53] 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 E1 should be read as indicating the existence (or absence) of gross departures from Zipf's Law, not as a precise test.

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