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

Jeremiah Dittmar*

June 28, 2010

Abstract

Zipf's Law characterizes city populations as obeying a distributional power law and is supposedly one of the most robust regularities in all of economics. This paper shows, to the contrary, that Zipf's Law only emerged in Europe between 1500 and 1800. It also shows that Zipf's Law emerged relatively slowly in Eastern Europe. The explanation I propose has two parts. First, because land and land-intensive intermediates entered city production as quasi-fixed factors, big cities were "too small" before 1500. Then, as trade and rising agricultural productivity relaxed the land constraint, it became possible for big cities to appear and Zipf's Law to emerge. Second, the institutions of the "second serfdom" in Eastern Europe were associated with delayed emergence. I find that laws limiting labor mobility and sectoral reallocation were associated with two factors that generate persistent deviations from Zipf's Law: relatively low variation in growth rates and a negative association between city sizes and growth rates ("non-random" growth). These legal institutions were also associated with the loss of several centuries of catch-up growth in Eastern European cities – a 1/3 reduction in city growth between 1500 and 1800. This institutionally-driven retardation has not previously been quantified. Taken together, these findings have important implications for how economists think about cities and, more broadly, economic growth.

*Contact: Department of Economics, American University, 4400 Massachusetts Avenue NW, Washington, DC 20016. Email: dittmar@american.edu. I thank Barry Eichengreen, Chad Jones, Christina Romer, Brad DeLong, Xavier Gabaix, Suresh Naidu, and seminar participants at UC Berkeley, UC Davis, Dartmouth, UBC, and American for comments 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 so exact and so "suspiciously like a universal law" as to be "spooky." Gabaix (1999a: 129) notes that it appears to hold in all economies and periods for which there are data. 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 benchmark 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 and examining its emergence leads us to ask important questions about cities and the determinants of growth.

This paper exploits data on the populations of European cities since 800 AD to examine and 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 geography matters but is not destiny.

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. As developments in trade and rising agricultural productivity relaxed the land constraint, it became possible for large cities to grow as fast as small cities. Zipf's Law emerged with this "modern" pattern of size independent growth 1500-1800, before the industrial revolution.

The paper also documents how institutional distortions shaped the pattern of city growth in Eastern Europe. In Eastern Europe, the legal institutions of the "second serfdom" placed severe restrictions on labor mobility and sectoral reallocation 1500-1800. These institutions were associated with persistent deviations from Zipf's Law and with a 1/3 reduction in city growth 1500-1800.

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). Figure 1 shows how Zipf's Law is apparent in data on city populations in three high income, three lower income, and three historical economies. 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]

2.2 Theories

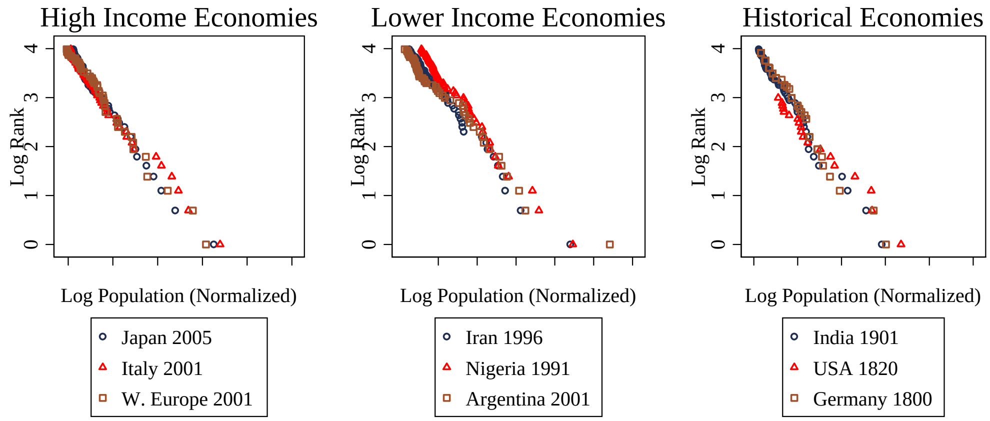
Two types of theories have been advanced to explain Zipf's Law. Random growth theories explain Zipf's Law as the outcome of a growth process in which all cities – big and small – draw growth rates from some common distribution. Geographic theories explain Zipf's Law as reflecting the distribution of natural advantages across locations.

[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.

Figure 1: 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.

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] Earlier contributions tying Zipf's Law to random growth include Krugman (1996a) and Simon (1955) and are reviewed in Gabaix (1999b, 2008).

Recent theoretical work has 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

[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.

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

[6]As discussed on p. 13, 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.

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, "Instead of city growth itself being random, it is fundamental economic characteristics of locations that are random."

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 et al. (2008). 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. Gabaix (1999b) observes a further anomaly: capital cities typically do not conform to Zipf's Law. I return to these points below.

Davis and Weinstein (2002) provide evidence in support of the locational fundamentals view. Davis and Weinstein find that a regional analogue to Zipf's Law holds 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.[7] 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. Similarly, 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."[8]

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

[8] 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. In addition, this paper examines data on all European cities while the de Vries data excludes much of Eastern Europe.

Archaelogical data confirm departures from Zipf's Law across a range of pre-modern or non-capitalist economies (Johnson 1980, Drennan and Peterson 2004, and Savage 1997).[9]

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 et al. (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 are intended to record (in thousands) the populations of urban agglomerations, not simply populations within administratively defined boundaries.[10] These data – henceforth the "Bairoch data" – are recorded every 100 years 800-1700, and then every 50 years to 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.

I test for measurement error in several ways. In Appendix B I compare the Bairoch data to the most comprehensive independent source for city population data, the database in de Vries (1984).[11] I document that the Bairoch and de Vries population figures are on average within 7 percentage points of each other and that there is no evidence of systematic shortfalls in the populations that the Bairoch data record for large cities.[12] However, it is possible that there is non-classical measurement error in both the Bairoch

[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]Bairoch et al. (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 was reviewed by 6 research institutes and 31 regional specialists in urban history.

[11]The Bairoch data covers all European cities that reached 5,000 inhabitants by or before 1800, has rich data from 1300 to 1850, and contains observations on 2,204 cities. The data in de Vries (1984) covers cities that reached a population of 10,000 between 1500 and 1800 and covers 379 cities.

[12]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.

Table 1: City Growth in Western Europe

Period	Cities	Mean	St. Dev.
(1)	(2)	(3)	(4)
800 - 900	31	0.09%	0.52%
900 - 1000	13	0.30%	0.77%
1000 - 1200	74	0.16%	0.30%
1200 - 1300	99	0.14%	0.58%
1300 - 1400	255	-0.22%	0.58%
1400 - 1500	187	0.06%	0.52%
1500 - 1600	321	0.18%	0.46%
1600 - 1700	514	-0.13%	0.55%
1700 - 1750	539	0.28%	0.60%
1750 - 1800	686	0.29%	0.63%
1800 - 1850	1,311	0.68%	0.78%

Note: Growth rates and standard deviations computed on an annualized basis.

data and de Vries (1984). In section 4.1, I show that the data would have to embody implausibly large non-classical measurement error for Zipf's Law to have actually held. In section 5, I show that the observed deviations from Zipf's Law are consistent with the narrative evidence collected by social historians.

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.

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 urban system.[13] 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 levels. These facts are documented in Appendix C.

Cities West of the Elbe River, which cuts through Eastern Germany, developed in a distinct institutional environment. The distinction between Western and non-Western

[13]See de Vries (1984), Nicholas (2003), Landes (1998), and Jones (1981).

cities was determined by two key aspects of economic institutions: first, the presence of institutions securing municipal autonomy for cities; second, the nature of the institutions determining the possibilities for mobility between the rural and urban sectors.[14] Town charters in the West typically guaranteed townspeople the right to legal proceedings in town courts, the right to sell their homes and to move, and freedom from most obligations associated with serfdom (e.g. arbitrary taxation, the provision of labor services, the head tax, and most forms of military service). These legal institutions fostered relatively secure property rights and the growth of urban commerce.[15] In Eastern and Ottoman-controlled Europe, cities developed in a different institutional environment. Under the Ottmans, cities were not granted municipal autonomy, allocations were more heavily influenced by administrative means, and city growth was also shaped by the "ruralization" of Christian populations.[16] In Eastern Europe, where German city law (*Deutsches Städtrecht*) had previously secured freedoms for townspeople and villagers, institutions limiting labor mobility and city autonomy were installed after 1500. These laws forbade seasonal migration, tied tenant farmers to rural estates, provided for the return of fugitive serfs, and limited the activities of merchants. An extensive literature documents the importance of the Elbe River as an institutional boundary distinguishing Western Europe from a central Eastern Europe in which the legal institutions of serfdom were strengthened after 1500.[17] Section 8 (below) examines the institutional geography, the nature of these laws, and their impact on city growth in greater detail.

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 key 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

[14] As discussed below, I follow the historical literature on the geography of economic instutions and take as Western all cities West of the Elbe and/or its tributary the Saale.

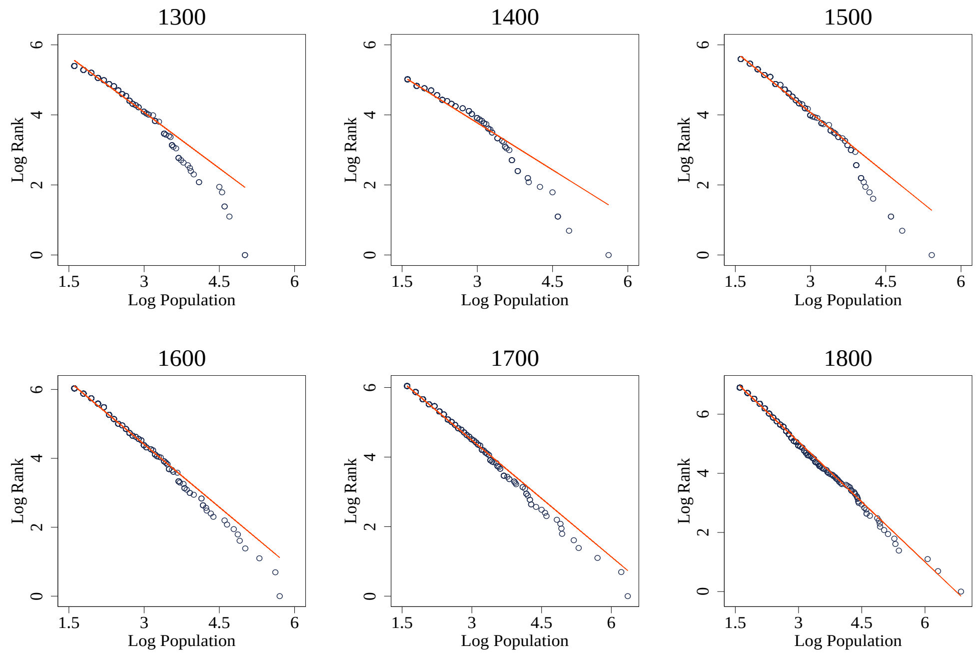
[15] See Pirenne (1927), Braudel (1979a, 1979b), Friedrichs (1995), Nicholas (2003), Scott (2005), Bideleux and Jeffries (2007), and van Zanden et al. (2010).

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

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

observed populations against fitted values associated with the robust non-parametric regression estimator proposed by Theil (1950).[18] 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.

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 et al. (1988).

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:

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

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.[19] Under the Gabaix test, we reject the null

[18] The way robust regression can be used to gauge departures from power laws is discussed below in section 4.1. 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 (2007) in terms of both small sample properties and precision.

[19] An earlier literature examined Zipf's Law with regressions: $\ln(R_i - 1/2) = \beta_0 + \beta_1 \ln S_i + \beta_2 \ln S_i^2 + \nu_i$. As discussed in Gabaix (2008), heteroskedasticity-robust standard errors will be biased down in

Table 2: Mean Square Deviations from Zipf's Law

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

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.

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

Year (1)	Observations (2)	δ_1 (3)	δ_2 (4)	Reject ZL (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 $S^* \equiv \text{cov}[(\ln S_i)^2, \ln S_i]/2\text{var}[\ln S_i]$. Following Gabaix (2008), 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.

---

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 over 1500-1800.

Quantile regression identifies more precisely where over the range of city sizes the curvature in the rank-size relation emerges.[20] Table 4 presents historical estimates of local, quantile slope parameters associated with equation (3). It shows that the big non-linearities were at the upper end of the city size distributions. In Table 4, as τ declines the estimates describe the local Zipf exponents (slopes) associated with progressively larger cities.[21] 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).

Table 4: Quantile Regression Estimates of Zipf Exponents

	Quantile Slope Parameters				
Year	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.

It is, however, conceivable that non-classical measurement error accounts for the pronounced deviations from Zipf's Law in the upper tail – that historical records undercount populations in big cities. It is also possible that historical data undercount populations of small cities. It bears noting that several characterizations of the deviations from Zipf's Law are observationally equivalent. We might observe either (1) that the big cities are too small and the small cities too few in number or (2) that there are too many mid-sized cities. As discussed below this pattern emerges when growth rates are negatively correlated with city populations.

[20]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 relaxes this assumption by assuming a piecewise linear loss function and minimizing the (asymmetric except in the case where $\tau = 0.5$) sum of absolute residuals. See Koenker (2005).

[21]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.

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. As shown in Appendix D, when data are generated by a stochastic power law, OLS estimators exhibit pronounced small sample bias. 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, which examination of OLS residuals typically cannot do.[22]

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. It shows that between 1500 and 1700 the biggest cities were consistently far smaller than they needed to be to satisfy a rank-size rule. For instance, the ten largest cities were on average less than 1/2 the size of the counter-factual Zipf-consistent populations in 1500. The magnitudes of the big city population shortfalls are so big that non-classical measurement error is not a plausible explanation for the observed deviations from Zipf's Law. The fact that the divergences are all shortfalls is consistent with the narrative evidence presented in the next section.[23]

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.

[22] See Koenker (2005), He et al. (1990), and Rouseeuw and Leroy (1987).

[23] When the data follow a distributional power law, ratios of city sizes in the upper tail have high standard deviations (see Gabaix 1999b). However, this variation is not likely to explain systematic, persistent short-falls in the upper tail. This can be verified with Monte Carlo simulations.

4.2 Implications

The observed deviations from Zipf's Law have three key implications. First, because 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, the fact that Zipf's Law emerged between 1500 and 1800 suggests that it is due to something beyond a power law distribution of propitious locations.

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.[24]

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. 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).

[24] Nicholas (2003: 7) observes that, "Probably no pre-modern city was as functionally specialised as modern industrial cities tend to be."

In early modern Europe, agricultural surpluses were limited, poor harvests brought famine, and secure access to food supplies was a precondition of the growth of large cities. The security of supplies was often dependent on the degree of control cities could exert on the surrounding countryside. For Paris, the largest city in 17th century Europe, the problem of securing foodstuffs was especially acute, and is repeatedly stressed by contemporary commentators. Generically, the largest cities faced similar challenges. 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. Cities on the Istrian and Dalmatian coast similarly controlled territories that stretched inland to the mountains. However, while a city's ability to control a rural district was typically contingent on the absence of a strong regional prince, urban territorial expansion was most often the result of purchases, foreclosed mortgages, and piecemeal treaty acquisitions – and not military conquest. In Germany, Nürnberg, Ulm, and Schwäbish Hall acquired hinterlands of 1,200, 830, and 330 km^2, respectively. The balance of political and economic influence might differ, but similar struggles emerged: Lübeck and Hamburg experienced a series of conflicts with the counts of Schleswig-Holstein and kings of Denmark over their rival claims on land, waterways, and resources. Broadly, growth at the upper end of the city size distribution was limited by the land constraint.[25]

If control over the countryside was important, transportation costs were the other 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.[26] As a result, cities preserved substantial forms of land-intensive production. There were gardens, fields, and areas devoted to livestock within cities themselves.[27] 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

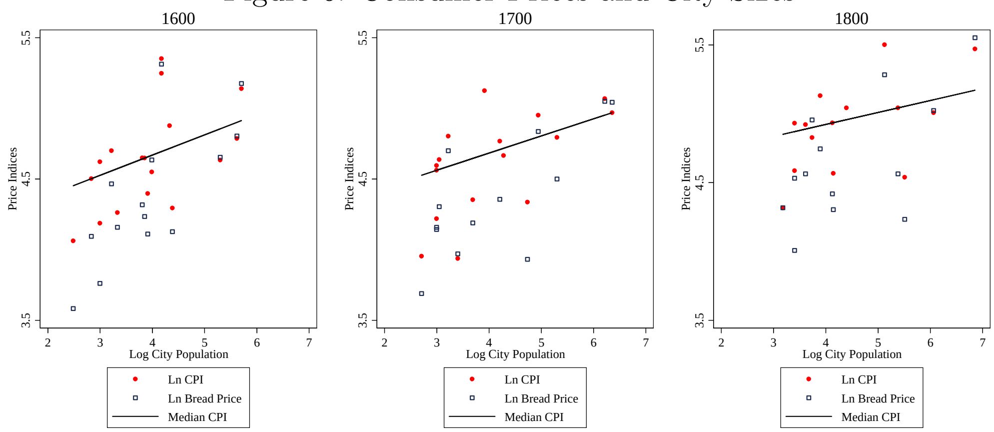
[25]See Scott (2004), de Vries (1976), Pounds (1990), Weber (1958), Pirenne (1958), Chittolini (1994), Livet (2003), Blockmans (1994), Nicholas (2003), Vilfran (1994), Miller (2008), Braudel (1979c).

[26]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. I discuss water-borne transport and estimate the growth advantage enjoyed by ports and cities on navigable rivers below.

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

(commercial hub of relatively urbanized Holland) than in Danzig (principal port of rural Poland). Pounds (1979: 61) observes that prices of agricultural products were broadly increasing in town size. The price data supports these observations. 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.[28] It shows that (1) consumer prices tracked bread prices, (2) prices were correlated with city size, and (3) that the correlation declined over time.[29] The data thus support the argument that while food prices were increasing in city size, the land constraint softened over time.[30]

Figure 6: Consumer Prices and City Sizes



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

Two factors relaxed the land constraint on city growth: the development of the international grain trade and increases in agricultural productivity. The grain trade was known to the Dutch as the *moedernegotie*: the "mother of all trades." In the Northern Netherlands, where city growth was unusually rapid, the bread and beer supply of city populations depended on grain imports. Grain came to be imported in quantities sufficient to feed 1 in 4 inhabitants of the Dutch Republic (van Tielhof 2002: 1). As shown in Table 6, 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." These developments were made possible by innovations in shipping technology (notably the introduction of the *fluyt* vessel) and an

[28] Allen (2001) provides data on the price of bread and consumer price indices in which basic land-intensive products account for two-thirds of indexed spending.

[29] OLS estimates confirm the decline in correlation between prices and city size. The OLS estimate (standard errors in parentheses) declines from 0.21 (0.08) in 1600 to 0.19 (0.07) in 1700 and 0.17 (0.07) in 1800.

[30] Historical evidence suggests that the price differentials associated with city size are not accounted for by the higher rents paid by bakers and other retail establishments selling basic wage goods. See Kriedte (1979) and Pounds (1979, 1990).

associated decline in freight shipping costs on maritime routes in the 16th century. 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).

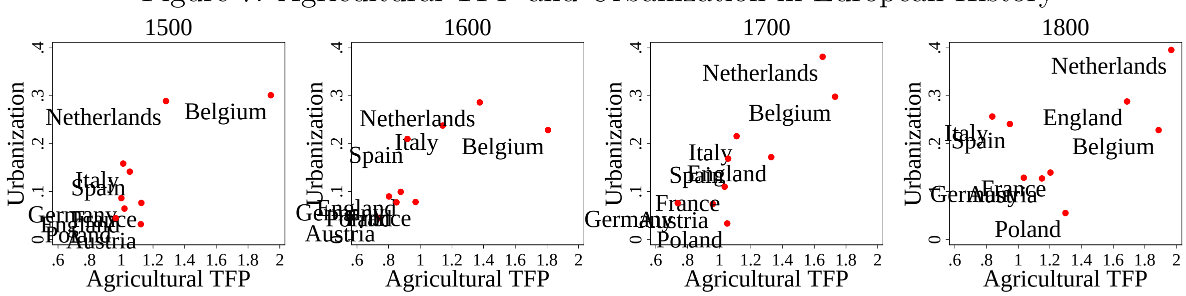
Table 6: Dutch Imports of Baltic Grain

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

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 in early modern Europe, and economies where city growth was concentrated also experienced relatively high rates of productivity growth in agriculture. Figure 7 shows the scatter plot of urbanization rates against Allen's (2003) estimates of agricultural total factor productivity in nine macroeconomies.

Figure 7: Agricultural TFP and Urbanization in European History



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

While the period 1500-1800 saw marked increases in agricultural productivity and in the trade in food, 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.[31] In general, mortality was increasing in city size. Other things equal, this limited the growth of large cities. However, the big revolutions in public health came after 1800 and mortality in large cities remained

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

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.

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.[32] 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 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.[33]

[32]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.

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

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

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

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

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

Production is Cobb-Douglas in technology (A), labor (young N^y and old N^o), and land (L)[34]:

$$Y_{it} = A_{it}(N_{it}^y)^\alpha (N_{it}^o)^\beta (L_{it})^{1-\alpha-\beta} \tag{7}$$

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

$$c_{it} = w_{it} = \frac{\partial Y_{it}}{\partial N_{it}^j} \quad j \in \{y, o\} \tag{8}$$

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.

6.3 Analysis of City Growth – The General Case

Individuals choose a city i subject to city-specific migration taxes τ_{it} and given the existing distribution of populations (wages). The individual maximization problem reduces to[36]:

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

[34]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).

[35]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 assumption corresponds to the situation in many Eastern European cities, which were owned by feudal lords. Alternately, one could assume that urban land is owned by a (small) patrician class. Introducing a class of urban landowners who 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.

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

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

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

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

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

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 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).[37]

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 assume that $A_{it} = A_t$. We now have the following variant of equation (13):

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

Here land has a positive income share because $\alpha+\beta < 1$. This fact secures the key feature of (15): city growth rates decline in population when land is fixed.[38] Broadly, one can imagine the long pre-modern era as one in which land entered 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 see in European city size distributions between 1300 and 1600. (See Appendix for a simple simulation.)

6.6 Case 3: Taxes and Distorted Growth

Prospective migrants face utilities that embody a tax τ_{it}.[39] The tax τ_{it} generates non-random growth when (and if) it falls hardest on larger cities. When the tax is constant across cities, it lowers the variance of city growth rates.

Assume land is absent from production and all cities have the same level of produc-

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

[38] This model is sketched in reduced form. The model as sketched has distinct regimes. A more detailed treatment could incorporate a transition that allows the share of income going to the fixed factor to decay.

[39] It is assumed in equations (**??**) and (**??**) that τ_{it} enters utility through the multiplicative amenity shock, but an additive structure would not change the story.

tivity (normalized to unity). Migration into the representative city is:

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

The associated growth rate is:

$$g_{it}^{N} = (1-\tau_{it})^{\frac{1}{1-\alpha}}\left(\frac{\alpha\epsilon_{it}}{u_t}\right)^{\frac{1}{1-\alpha}} - \delta \tag{17}$$

τ_{it} captures the serfdom distortion. It can be thought of as embodying the effects of migration restrictions and legislation imposing higher effective taxes on larger cities (via a productivity distortion) – as generating a negative correlation between size and growth.

The institutions of the second serfdom had such biases. Historically, the share of economic activity accounted for by merchants and capitalists involved in the coordination and financing of longer-distance trade was increasing in city size. An institutional set-up biased against these activities is one with higher effective taxes on big cities. In Poland, merchants were in 1565 forbidden from owning land, travelling abroad, and engaging in international trade. In Prussia and Poland-Lithuania, land owners won the right to export their produce directly, circumventing local cities and merchants, and without paying otherwise required export taxes. In Poland, Prussia, and Bohemia price maxima were also placed on urban goods tilting the terms of trade towards agricultural landowners, lowering city incomes and the incentive to migration.[40] In Hungary, the legislation passed after the defeated peasant uprising of 1514 limited the mobility of tenant farmers and eliminated the legal autonomy of towns. Migration restrictions also appear to have hit larger towns hardest. Larger cities typically had to draw migrants from relatively far afield. However, peasant migrants fleeing serfdom were not able to safely travel great distances. In the late 1600s the Austrians sent an emissary to Krakow to press the Polish authorities to implement their treaty agreement and return fugitive Silesian serfs. While this suggests instances of remarkable mobility, Wright (1961) observes that for most Bohemian serfs Poland was too distant to be an attainable asylum.

It is important observe that even a tax that does not vary with city size will delay the emergence of Zipf's Law by lowering the variance of the city growth rate.[41] Assume that τ is constant across cities and for simplicity denote the variance of $\epsilon_{it}^{1/(1-\alpha)}$ with $\sigma_{\tilde{\epsilon}_t}^2$.

[40] See Blum (1957), Carsten (1954), and Kula (1962).

[41] Gabaix (1999b) discusses how a growth process with a higher variance leads to relatively speedy emergence of Zipf's Law.

From equation (17), the variance of the growth rate is declining in τ:

$$\sigma^2_{g^N} = (1 - \tau_t)^{\frac{2}{1-\alpha}} \left(\frac{\alpha}{u_t}\right)^{\frac{2}{1-\alpha}} \sigma^2_{\tilde{\epsilon}_t} \tag{18}$$

Moreover, these results hold when there is no distortion in amenities, but discriminatory institutions operate such that productivity is deflated by τ_{it}:

$$Y_{it} = (1 - \tau_{it})A_{it}(N^y_{it})^{\alpha}(N^o_{it})^{\beta}(L_{it})^{1-\alpha-\beta} \tag{19}$$

In this case, the city growth rate g^N_{it} suffers from a distortion identical to the one generated by a direct tax on migration. City growth is again negatively correlated with city population when distortions rise in city size, and the variance of the growth rate is depressed even when they do not.

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

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

Table 9 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. From 1500 forwards there is no significant correlation between size and growth.

Next I group cities into size quantiles and examine the distribution of growth rates within each quantiles. Figure 8 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 9 presents data for 1200-1800 and shows how random growth emerged from 1500 forward.

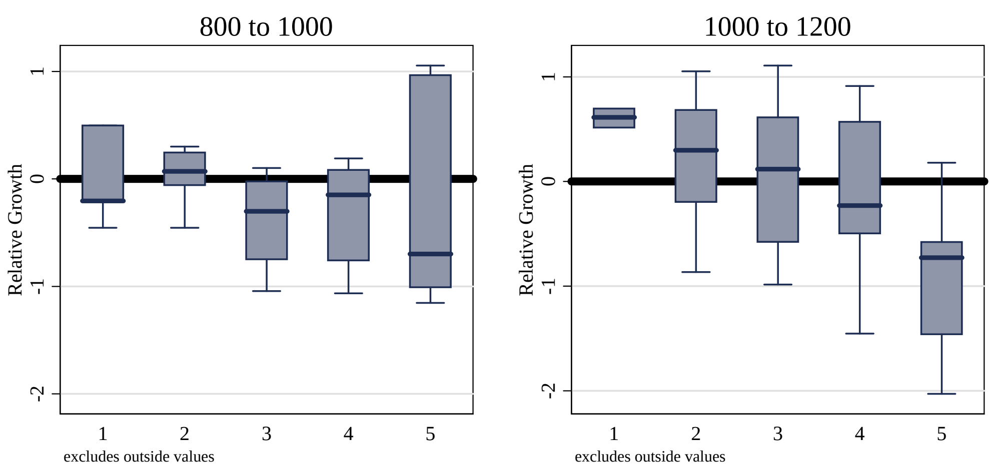
Figure 10 compares the distribution of growth rates for the top 10 percent and bottom 90 percent of cities using nonparametric kernel densities. It shows that up through the period 1500-1600, the largest cities consistently grew more slowly than smaller cities. It also shows that by 1700-1800, large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

Table 9: 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 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.

Figure 8: City Growth Rates By Population Quintile 800-1200



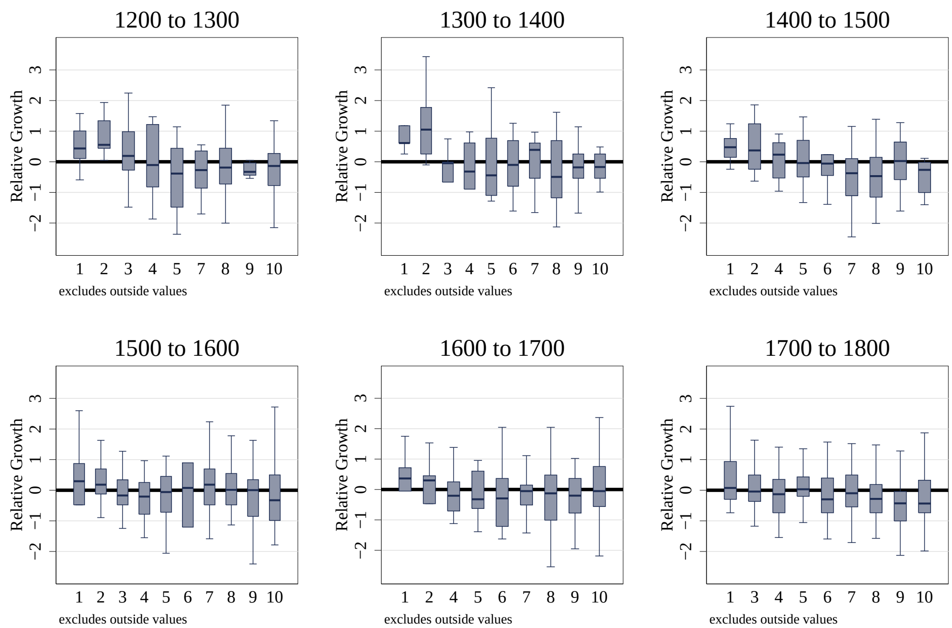
Note: This figure documents how normalized growth rates varied with city size. The smallest cities are in quintile 1, the largest in quintile 5. The boxes describe the interquartile range. The line within each box is the decile's median growth rate. The "whiskers" mark the adjacent values.

approximately identical distributions.

7.2 Locational Fundamentals

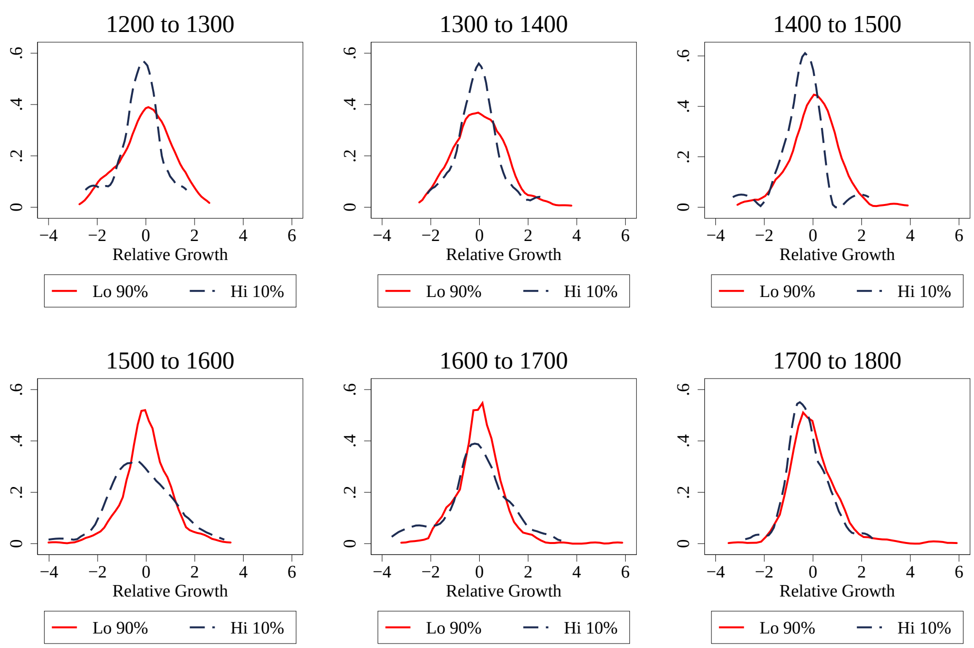
Geographic theories of Zipf's Law hold that the distribution of natural advantages across locations is the underlying determinant of city population distributions. This paper documents that European city populations have not always obeyed a power law distribution.

Figure 9: 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.

Figure 10: The Distribution of City Growth Rates

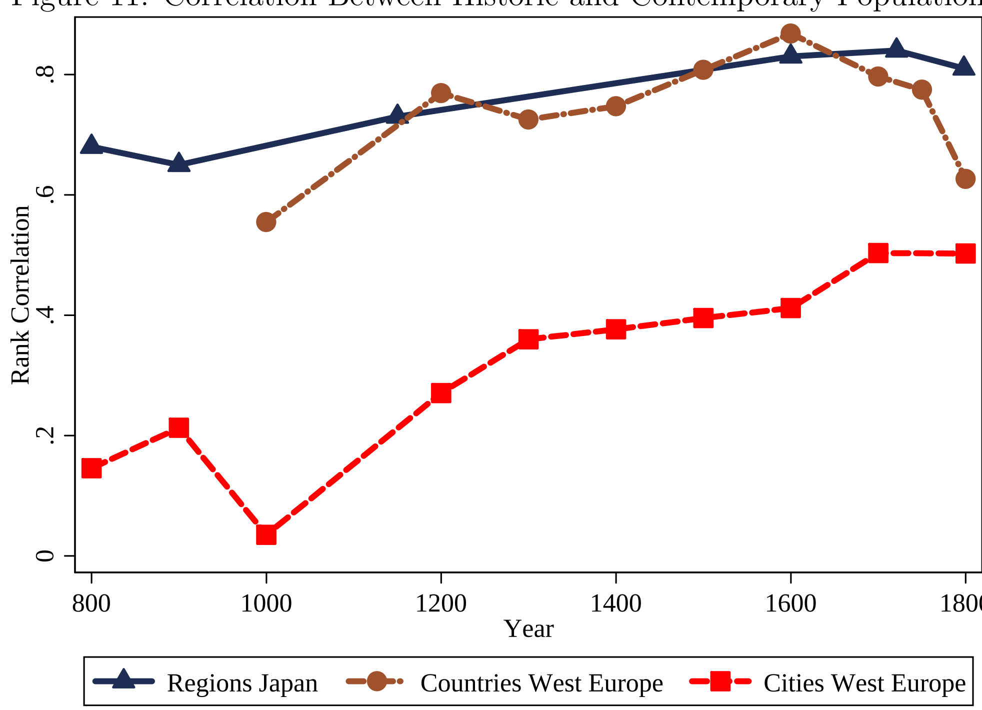


Note: This figure presents kernel densities of the distribution of growth rates for the largest 10 percent (Hi 10%)and smallest 90 percent (Lo 90%) of cities.

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. In other instances, locational advantages may be quite literally constructed (harbors may be dredged, important canals dug, etc.).

Two key findings have been used to support a geographic theory of Zipf's Law. The first is Davis and Weinstein's (2002) observation that a regional analogue of Zipf's Law holds over many centuries in the Japanese data and that Japanese regional population densities in the past are highly correlated with contemporary population densities. 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 Davis and Weinstein find 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 a distinction should be drawn between regional and city-level population data and that discussions of Zipf's Law should (data permitting) focus on cities.

Figure 11: Correlation Between Historic and Contemporary Populations



City population data are from Brinkhoff (2008) and Bairoch et al. (1988). National population data are from Eurostat and Acemoglu et al. (2005). Correlations for regional and national data are calculated using population densities. Correlations for Japanese regions are from Davis and Weinstein (2002).

The second key finding for geographic theories is that city size hierarchies have been remarkably stable in the 20th century even in the face of large temporary shocks. Davis and Weinstein (2002) and Boosker et al. (2008) 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. However, the long run evidence suggests that European city size hierarchies were relatively dynamic.

The historical data reveal significant churning in European city hierarchies over the course of centuries. The data thus 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 Europe's largest cities were concentrated in Northwestern Europe. There were also sharp shifts in urban populations at more local levels. 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. Cologne was the largest German city between 1200 and 1500; today it is the seventh largest. Augsburg went from being the largest German city in 1600 to 8th largest in 1800 and the 24th largest in 2006. 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 (population 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 itself 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.[42]

8 City Populations with Big Distortions: The Case of Eastern Europe

In this subsection I explain how the institutions of the second serfdom distorted city growth in Eastern Europe and were associated with deviations from Zipf's Law.

The central institutions of the second serfdom restricted the free ciculation of labor (Makkai 1979: 235). Bideleux and Jeffries (2007: 161) observe that that, "the legislative

[42]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).

strengthening of eastern European serfdom had begun during the 1490s" and that the new laws reflected, "concerted action to restrict the rights and mobility of the peasantry."[43]

After 1500, new institutions limiting labor mobility and the autonomy of cities were installed in the economies East of the Elbe River.[44] The laws passed in central Eastern Europe typically required peasants entering towns to carry proof that they had obtained their landlord's permission to travel and formalized systems of adscription that legally bound tenant farmers to rural estates. The laws provided for the cross-border return of fugitive serfs, limited the activities of merchants and restricted the privileges of towns. The new institutions rolled back previously guaranteed legal freedoms and end ed institutional convergence between Eastern and Western Europe.[45]

I construct an index of second serfdom laws, determining whether a given city was located in a polity with the legal restrictions on peasant mobility that were the institutional heart of the second serfdom. **Table 9** records the dates of passage of the principal laws limiting the mobility of tenant farmers. For a given city, the serfdom index captures the presence of local laws limiting labor mobility, from the dates given in Table 9 until the date of the first emancipation decree issued in the territory. **Table 10** provides a complete list of emancipation decrees.[46]

Regression analysis confirms that the institutions of the second serfdom were associated with large variations in city growth. In this section I present results from a baseline model that includes controls for initial population, regional fixed effects, and period fixed effects.[47] I present results with and without controlling for the growth effects associated with political primacy.[48] The key finding is that the institutions of the second serfdom

[43]See also Samsonowicz and Mączak (1985), Bogucka (1982), Pachs (1994), Berend (1986), Süchs (1988), Kahan (1973), Anderson (1974a, 1974b), Wright (1958, 1975), Hagen (1998), Melton (1988), Blum (1957, 1978), Carsten (1954), Brenner (1974), and Topolski (1982, 1974).

[44]Historians come to a striking consensus on the Elbe boundary. Kriedte (1979: 22) notes that the Elbe River "became the most significant socio-economic divide in Europe." Berend (1986: 333-334) observes that, "with astonishing precision," the Elbe constituted, "The sharp line of demarcation between the economic and social structures that divided Europe in two after approximately the year 1500." Robisheaux (1998: 111) observes that, "The line runs South along the Elbe river, through Saxony, and along the Erzebirge Mountains and heavily forested border between Bohemia and Bavaria." See also Bideleux and Jeffries (2007), , Süchs (1988), Maddalena (1977), and Brenner (1974).

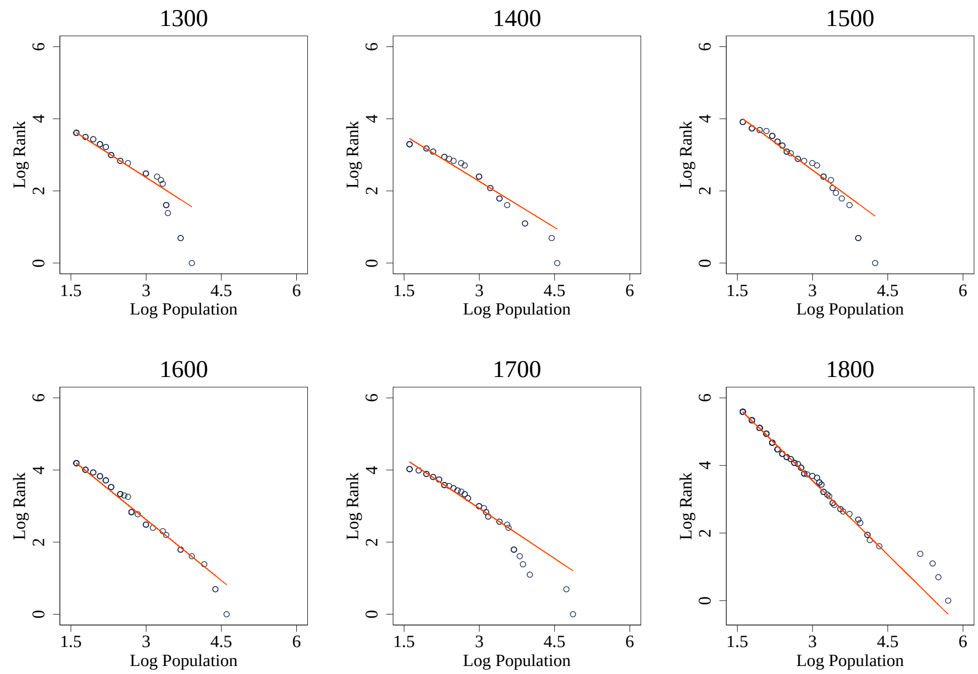
[45]These laws were instituted in economies where cities were walled and entered only through guarded gates (Friedrichs (1995: 21-22). Feudal lords often maintained control over the gates and employed the gatekeepers (Nicholas 2003). See also Miller (2008: 37) and Mols (1955: 347-348). The penalties associated with illegal movement were severe. The Prussian legal ordinances of 1494 stipulated that runaway peasants could be hanged by their masters without trial or arbitration. Carsten (1954: 108).

[46]The results presented below reflect a coding where SERF=1 over periods when legal restrictions were in place a majority of the time. A coding that relies on a continuous measure of fraction of years under serfdom yields similar results.

[47]Adding country fixed effects has no substantive impact on the estimated associations between growth and the institutions of the second serfdom.

[48]Gabaix (1999b) observes that capitals typically do not conform to Zipf's Law. As observed on p. **??**,

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



Note: This figure plots raw data on city populations (S_i) and size rankings (R_i), and fitted values estimated using robust non-parametric Theil regression and the model: $\ln(R_i) = \alpha - \beta \ln(S_i) + \epsilon_i$. See note to Figure 2.

Table X: Comparison of City Growth & Deviations from Zipf's Law

	Correlation Between Size & Growth		Deviation from Zipf's Law	
Period Starting	Western Europe	Eastern Europe	Western Europe	Eastern Europe
(1)	(2)	(3)	(4)	(5)
1400	-0.27 **	-0.38 *	4.21%	4.31%
1500	-0.05	-0.39 **	1.51%	3.73%
1600	0.00	-0.29 *	0.58%	1.32%
1700	-0.07	-0.03	0.50%	4.89%
1800	--	--	0.18%	0.85%

Note: .

were associated with a 1/3 cut in city growth.

The baseline estimating equation examines the association between growth and the

Berlin and Vienna were unusual in growing quickly and being substantially larger than predicted by robust regression in 1800.

Table X: Dates of Principal Legal Restrictions on Free Migration

Historic Territory (1)	Contemporary Location (2)	Date (3)
Austria	Austria	1539
Bohemia	Czech Republic	1487
Brandenburg	Eastern Germany	1528
Hungary	Hungary	1514
Livonia	Estonia & Latvia	1561
Mecklenberg	Northeastern Germany	1654
Poland	Poland	1495
Pomerania	Northeastern Germany	1616
Prussia	Eastern Germany, Poland	1526
Romanian Wallachia	Romania	late 1500s
Russia	Russia	1640s/1700s
Saxony	Eastern Central Germany	--
Schleswig-Holstein	Northern Germany	1617
Silesia	Czech Rep., Poland, East Germany	1528

Note: See Appendix for sources.

Table X: Dates of Emancipation Decrees in Eastern Europe

Historic Territory (1)	Year (2)
Poland (Grand Duchy of Warsaw), Prussia	1807
Estonia	1816
Courland	1817
Livonia	1819
Mecklenberg	1820
Saxe-Altenberg	1831
Saxony	1832
Scwarzburg-Sonderhausen, Reuss (older), Saxe-Weimar, Austria, Saxe-Gotha, Anhalt-Dessau-Köthen	1848
Schawrzburg-Rudolstat,Anhalt-Bernberg	1849
Saxe-Meiningen	1850
Reuss (younger)	1852
Hungary	1853

Note: See Appendix for sources.

laws of the second serfdom:

$$\log \text{growth}_{i,t} = \alpha + \beta(\log \text{size})_{i,t} + \sum_j \gamma_j \text{region}_j + \sum_k \eta_k \text{year}_k + \theta(\text{second serfdom})_{i,t} + \epsilon_{i,t} \tag{20}$$

Table 11 presents results for several geographic samples.[49] Table 11 begins with a sample

[49]The non-Russian sample includes Kaliningrad (Königsberg) and cities in the Baltics. The Broad Central Europe sample comprises the cities in contemporary Germany, Austria, Poland, Hungary, the

Table 11: Baseline Analysis of City Growth From 1300 to 1850
Dependent Variable is Log City Growth

Independent Variable	Non-Russian Europe	Broad Central Europe	Central Eastern Europe	Non-Russian Europe	Broad Central Europe	Central Eastern Europe
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log Size	-0.01	0.01	0.02	-0.03 **	-0.01	-0.02
	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)
Serf	-0.10 **	-0.17 **	-0.14 **	-0.10 **	-0.15 **	-0.10 **
	(0.04)	(0.05)	(0.05)	(0.03)	(0.04)	(0.04)
West	0.05	-0.16 **	-0.13 **	0.04	-0.17 **	-0.12 **
	(0.05)	(0.03)	(0.03)	(0.05)	(0.02)	(0.03)
East	0.19 **			0.19 **		
	(0.05)			(0.05)		
Capital				0.42 **	0.41 **	0.49 **
				(0.05)	(0.06)	(0.05)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4,069	1,523	773	4,069	1,523	773
F Statistic	65.06	28.60	20.97	65.32	30.74	35.85
R Squared	0.14	0.16	0.23	0.16	0.17	0.25
SE Clustered	On City	On City	On City	On City	On City	On City

Note: Non-Russian Europe includes cities in Ottoman Europe and the Baltics. Significance with 90 and 95 percent confidence denoted with "*" and "**", respectively.

of non-Russian cities. However, given the emphasis Acemoglu et al. (2005) place on the role Atlantic trade played in early modern city growth, it makes sense to also examine samples that exclude Atlantic cities and focus on the impact of serfdom on city growth within central Europe. Across geographic samples two parameter estimates stand out. First, there is a strong *positive* association between city growth and location in Eastern Europe: an Eastern location is associated with an increase of over 0.1 log points of growth every 100 years (at least 10 extra percentage points).[50] As shown below, this association is robust and appears to reflect the catch-up growth advantage held by the relatively small cities of the East. Second, there is a significant negative association between serfdom and growth roughly equal to the positive association between Eastern location and growth. Among Central Eastern European cities, serfdom is associated with a decline in growth

former Czechoslovakia, and France. The Central Eastern Europe sample is limited to cities in Germany, Austria, Poland, Hungary, and the former Czechoslovakia.

[50]In samples restricted to Western and Eastern Europe, a Western location has a negative association with growth.

of 0.14 log points. Among non-Russian cities the estimated decline is 0.10 log points.[51]

These magnitudes are economically significant. Over 100 years, a decline of 15 percentage points implies a decline of 0.0016 in the annual growth rate. But mean growth in the East from 1300 to 1850 was 0.0054 and mean annual growth in the East was not more than 0.0027 before 1750. For illustration: $(0.0016)/(0.0016+0.0054) \approx 23$ percent, and $(0.0016)/(0.0016 + 0.0027) \approx 38$ percent. Parameter estimates of these magnitudes suggest that the imposition of laws restricting labor mobility may have cut Eastern city growth by two fifths over several centuries.

In every sample, the positive association between an Eastern location and growth and the negative association between serfdom and growth essentially cancel each other out. As discussed above, overall growth rates in Western and Eastern cities are roughly comparable. The fact that Eastern cities grew relatively quickly, but that serfdom was associated with slow growth suggests that the institutional framework may have prevented or delayed a catch-up process otherwise under way.

As observed above (section 6.6), we expect a tax that depresses growth to reduce the variance of city growth rates. This matters because Zipf's Law emerges relatively slowly where the variance of growth rates is low. **Table 12** compares similarly sized cities and shows that the institutions of the second serfdom were associated with relatively low variance in city growth between 1500 and 1700. It further shows that from 1500 to 1800 the coefficients of variation for the largest cities (population at least 50,000) were lower in cities exposed to serfdom than in their Western counterparts.

The institutions of the second serfdom may also have delayed the emergence of Zipf's Law by generating non-random growth. As discussed above, political capitals were unusual in being large and fast growing cities. When they are excluded from the analysis, we observe (i) a negative correlation between growth and size across institutional regimes, and (ii) that the negative correlation is both substantially stronger and more imprecisely estimated for cities exposed to serfdom. **Figure 6** illustrates this by pooling normalized data and plotting normalized growth rates against normalized city sizes.

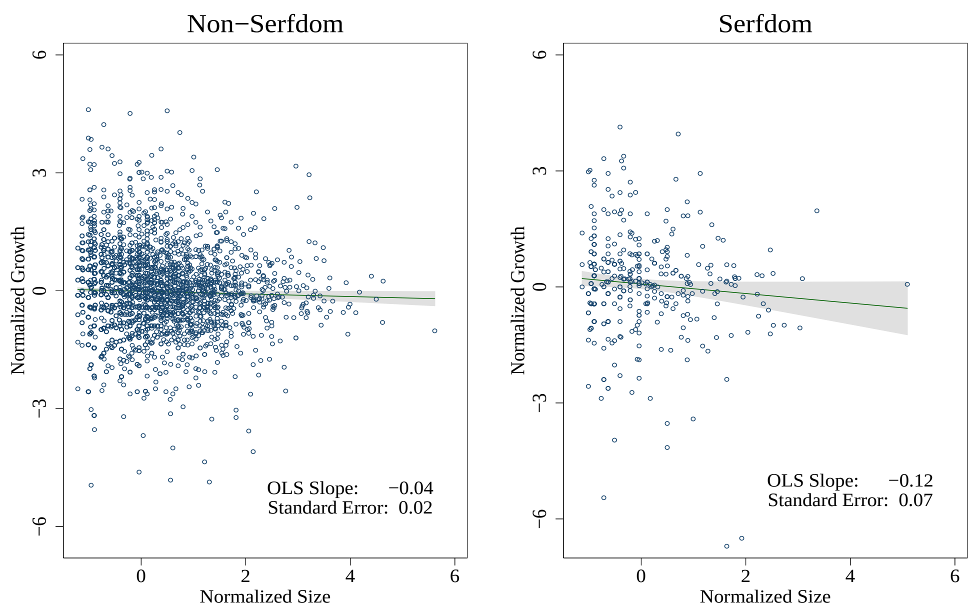
[51] Russian cities are excluded for two principal reasons. First, data on Russian city populations is uniquely noisy (Bairoch 1988: 170). Second, given the distinct nature of Russian institutions, historians distinguish Russia from the states that fall within the original limits of the second serfdom (Makkai 1975: 233).

Table 12: The Second Serfdom and the Variance of City Growth Rates

Period Starting	Institutional Regime	Coefficients of Variation for Cities Grouped by Population					
		5k - 6k	7k - 9k	10k-14k	15k - 24k	25k - 49k	50k +
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1500	Serfdom	0.6	0.1	0.4	4.9	2.3	0.0
	No Serfdom	1.2	4.9	2.2	2.2	9.0	4.7
	Ratio	0.5	0.0	0.2	2.2	0.2	0.0
1600	Serfdom	2.2	2.0	2.1	2.1	2.2	6.1
	No Serfdom	4.7	2.4	4.3	9.8	4.7	37.1
	Ratio	0.5	0.8	0.5	0.2	0.5	0.2
1700	Serfdom	1.8	2.6	1.3	11.1	5.0	1.4
	No Serfdom	1.8	1.8	1.9	5.3	3.0	4.5
	Ratio	1.0	1.4	0.7	2.1	1.7	0.3

Note: This table presents coefficients of variation for cities grouped by population. For instance, cities in column (3) had populations between 5 and 6 thousand, inclusive.

Figure 6: Growth and City Size



Note: This graph shows the relationship between growth and size in pooled, normalized data. The data includes all cities that were not political capitals.

9 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 and city growth rates became random, in the sense of being independent of city population.

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, however, consistent with theories emphasizing random growth in the emergence of Zipf's Law.

References

[1] Acemoglu, D. et al. (2005), "The Rise of Europe," *American Economic Review*, Vol. 95, No. 3, pp. 546-579.

[2] Acemoglu, D. et al. (2002), "The Rise of Europe," NBER Working Paper 9378. Vol. 110, pp. 195-227.

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

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

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

[7] Aubin, H. (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.

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

[9] Bairoch, P. et al. (1988), *La Population Des Villes Européenes*, Geneva; Droz.

[10] Bath, S. 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.

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

[12] Bideleux, R. and I. Jeffries (2007), *A History of Eastern Europe*, London; Routledge.

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

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

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

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

[17] Bogucka, M. (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.

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

[19] Brakman, S. et al. (2004), "The strategic bombing of German cities during WWII and its impact on city growth," *Journal of Economic Geography*, Vol. 4, pp. 201-218.

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

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

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

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

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

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

[26] Cook, C. and J. Stevenson (1978), *Longman Atlas of Modern British History*, London; Longman.

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

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

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

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

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

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

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

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

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

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

[37] Gabaix, X. and R. Ibragimov (2007), "Rank-1/2: A Simple Way to Improve the OLS Estimation of Tail Exponents," NBER Technical Working Paper 342.

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

[39] Hagen, W. (1998), "Village Life in East-Elbian Germany and Poland, 1400-1800," in T. Scott ed., *The Peasantries of Europe from the Fourteenth to the Eighteenth Centuries*, London; Longman.

[40] He, X. et al. (1990), "Tail Behavior of Regression Estimators and their Breakdown Points," *Econometrica*, Vol. 58, No. 5, pp. 1195-1214.

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

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

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

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

[45] Hollander, M. and D. Wolfe (1999), *Nonparametric Statistical Methods*, 2nd edition, New York; Wiley and Sons.

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

[47] Jedin, H. et al. (1970), *Atlas Zur Kirchen-Geschichte*, Freiburg; Herder.

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

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

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

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

[52] Klima, A. (1979), "Agrarian Class Structure and Economic Development in Pre-Industrial Bohemia," *Past and Present*, No. 85, pp. 49-67.

[53] Kochanowicz (1989), "The Polish Economy and the Evolution of Dependency," in D. Chirot ed., *The Origins of Backwardness in Eastern Europe"*, Berkeley; University of California.

[54] Koenker, R. (2005), *Quantile Regression*, Cambridge; Cambridge University.

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

[56] Krugman, P. (1981), "Trade, Accumulation, and Uneven Development," *Journal of Development Economics*, Vol. 8, pp. 149-161.

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

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

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

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

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

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

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

[64] Marshall, A. (1920), *Principles of Economics*, 8th Ed., London; MacMillan.

[65] Meigs, R. (1973), *Roman Ostia*, Oxford; Clarendon.

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

[67] Miller, J. (2008), *Urban Societies in East-central Europe, 1500-1700*, London; Ashgate.

[68] Mols, R. (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.

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

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

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

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

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

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

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

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

[77] Rousseeuw, P. and A. Leroy (1987), *Robust Regression and Outlier Detection*, New York; John Wiley.

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

[79] Samsonowicz, H. and A. Maçzak (1985), "Feudalism and capitalism: a balance of changes in East-Central Europe," in P. Burke ed., *East-Central Europe in transition*, Cambridge; Cambridge University.

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

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

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

[83] Smith, A. (1776), *The Wealth of Nations*, 5th Ed. [1904], London; Methuen.

[84] Smith, C. (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.

[85] Smith, C. (1982), “Modern and premodern urban primacy,” *Comparative Urban Research*, Vol. 11, pp. 79-96.

[86] Soo, K. (2005), “Zipf’s Law for cities: a cross-country investigation,” *Regional Science and Urban Economics*, Vol. 35, pp. 239-263.

[87] Stillwell, R. et al. (1976), *Princeton Encyclopedia of Classical Sites*, Princeton; Princeton University.

[88] Süchs, J. (1988), “Three Historical Regions of Europe,” in J. Keane ed., *Civil Society and the State*, London; Verso.

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

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

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

[92] Todorov, N. (1983), *The Balkan City, 1400-1900*, Seattle; University of Washington Press.

[93] Topolski, J. (1994), *The Manorial Economy in Early-Modern East-Central Europe*, Aldershot; Variorum.

[94] Topolski, J. (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.

[95] Topolski, J. (1974), “Economic Decline in Poland from the Sixteenth to the Eighteenth Centuries,” in P. Earle ed., *Essays in European Economic History*. Oxford; Clarendon.

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

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

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

[99] de Vries, J. (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.

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

[101] Wallerstein, I. (1974), *The Modern World-System: Capitalist Agriculture and the Origins of the European World-Economy in the Sixteenth Century*, New York; Academic Press.

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

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

[104] Wright, W. (1975), "Neo-Serfdom in Bohemia," *Slavic Review*, Vol. 34, No. 2, pp. 239-252.

[105] Wunder, H. (1978), "Peasant Organization and Class Conflict in Eastern and Western Germnay," *Past and Present*, No. 78 (February), pp. 47-55.

[106] Zientara, B. (1982), "*Melioratio terrae*: the thirteenth-century breakthrough in Polish history," in J. Fedorowicz et al. eds., *A Republic of Nobles: Studies in Polish History to 1864*, Cambridge; Cambridge University.

— Not For Publication —

A Appendix: Data

City populations are from Bairoch et al. (1988) and de Vries (1984). City locations are from Bairoch et al. (1988), cross-checked using http://www.batchgeocode.com/.

Data on the dates and and nature of the laws restricting labor mobility and limiting the rights of urban groups under the second serfdom is from Makkai (1975), Topolski (1982), Blum (1957), Carsten (1954), Szelényi (2006), Davies (1981), Pachs (1994), Hellie (1971), Kahan (1973), Kamiński (1975), Bogucka (1984), Melton (1988), Maddalena (1977), and Bideleux and Jeffries (2007). Data on the dates of emancipation decrees is from Blum (1978). The historical coding of the Polity-IV index of constraints on arbitrary executive authority is from Acemoglu et al. (2002, 2005). DeLong and Shleifer (1993) class regional institutions as either promoting relatively unrestrained and autocratic rule ("prince") or as securing relative freedom ("free"). I extend this coding to Poland and Ottoman Europe, neither of which meet the criteria for classification as "free" between 1300 and 1850 (this is confirmed by DeLong).

Data on the historical location of universities are from Darby (1970), Jedin (1970), and Bideleux and Jeffries (2007). Data on the historical location of religious institutions are from Magosci (1993) and Jedin (1970). Data on Roman settlements are from Stillwell et al. (1976).

Data on the historical location of ports are from Acemoglu et al. (2005), supplemented by data in Magosci (1993) and Stillwell et al. (1976), and the sources cited in section **??**. The data in this paper supplements Acemoglu et al. (2005) by coding for cities that were historically ports on the Baltic. These cities include: St. Petersburg, Gdańsk, Kaliningrad, Szczezin, Rostock, and Lübeck. In addition, the coding in this paper accounts for Mediterranean and Black Sea ports omitted in Acemoglu et al. (2005): Gaeta, Fano, Kerch, Korinthos, Pozzuoli, and Trapani.

Data on the location of navigable rivers are drawn from Magosci (1993), Pounds (1979, 1990), Livet (2003), Cook and Stevenson (1978), Graham (1979), Stillwell et al. (1976), and de Vries and van der Woude (1997). The coding captures the principal historically navigable waterways, and does not class as "navigable" waterways that required substantial improvements (dredging, re-channeling, etc.) and became navigable only over the early modern era.

National-level urbanization rates are from Acemoglu et al. (2005). Data on national-level agricultural TFP are from Allen (2003). Data on consumer prices in European cities are from Allen (2001).

B Appendix: Measurement Error

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

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 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

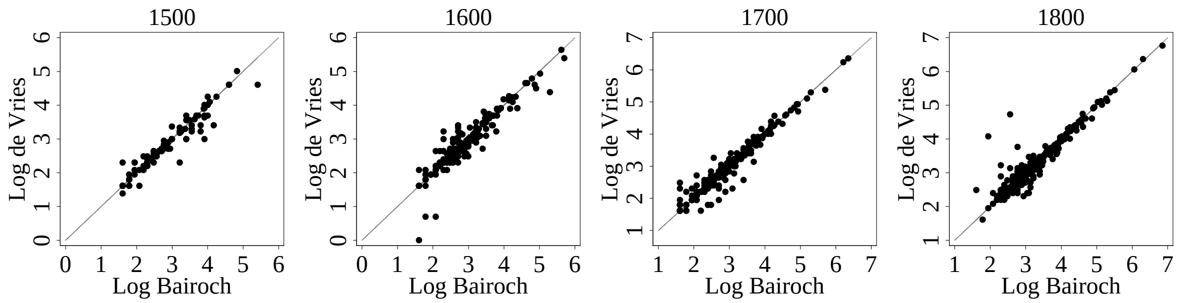
Note: This table compares population data from Bairoch et al. (1988) and de Vries (1984). Column (3) presents the correlation between recorded values. Columns (4) to (8) examine the ratio of these values.

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.[52] However, it is possible that

[52]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

there is non-classical measurement error in both the Bairoch data and de Vries (1984). In section 4.1, I show that the data would have to embody implausibly large non-classical measurement error for Zipf's Law to have actually held. In section 5, I show that the observed deviations are consistent with the narrative evidence.

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.

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 5 shows how Zipf's Law emerged between 1400 and 1800 in the six leading national economies of Western Europe.

D Appendix: Small-Sample Estimators for Zipf Exponents

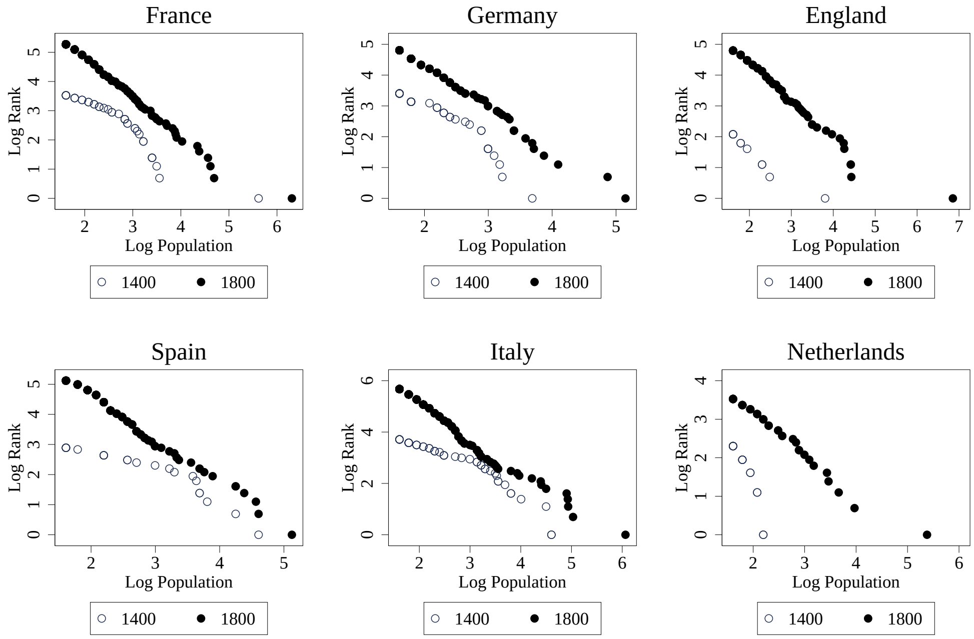
This appendix discusses the estimation of Zipf exponents and some properties of the Theil estimator.

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

$$\ln R_i = \alpha - \beta \ln S_i + \epsilon_i \tag{21}$$

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.

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



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.

$$\ln(R_i - 1/2) = \alpha - \beta \ln S_i + \epsilon_i \tag{22}$$

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.[53] However, as Gabaix and Ioannides (2004) observe, the small sample

[53]See Soo (2005), Newman (2005), and Clauset et al. (2007). For a sample of n cities with sizes S_i ordered so that $S_{(1)} \geq \ldots \geq S_{(n)}$, the Hill estimator is: $\hat{\beta}_H = (n-1)/\sum_{i=1}^{n} \left[\ln(S_{(i)}) - \ln(S_{(i+1)})\right]$.

biases associated with the Hill estimator can be quite high and very worrisome. Moreover, the Hill estimator is the MLE under the null hypothesis that the data generating process is a distributional (and specifically Pareto) power law, but is not the MLE if the empirical distribution is not Pareto. For these reasons, this paper does not present estimates using the Hill estimator.

Robust regression techniques have been designed for situations where sample sizes are small and/or outliers may have an undue impact on OLS estimates. A number of robust regression estimators use the framework provided by the median. In particular, the nonparametric estimator derived from Theil (1950) is intuitive, asymptotically unbiased, robust with small samples, allows us to go some distance in addressing the problem posed by outliers, and has not been exploited in the Zipf's Law literature.[54] 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.

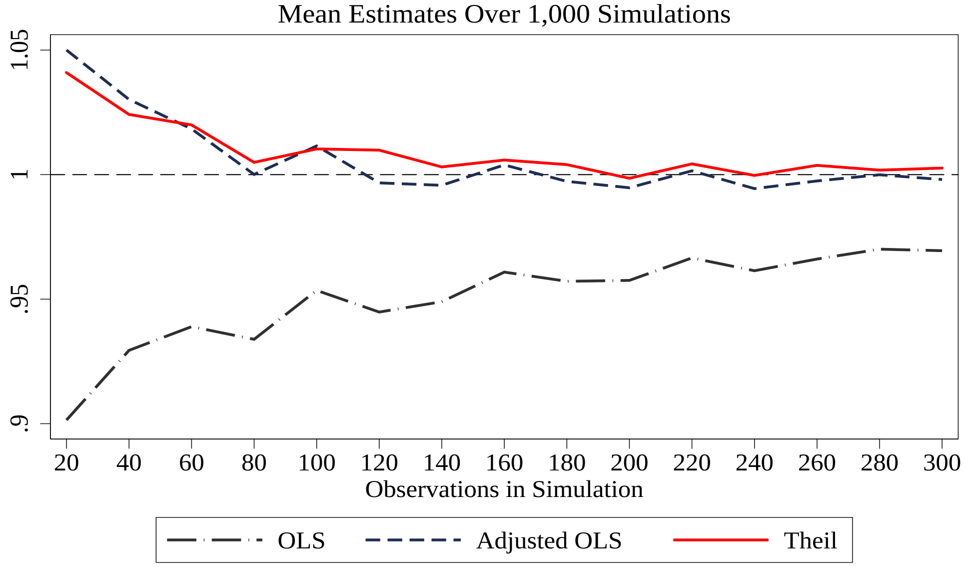
The Theil estimator is competitive with the rank-adjusted OLS estimator suggested in Gabaix and Ibragimov (2007) in eliminating small sample bias. This is evident in **Figure B1**, which uses simulated data (generated by a process with Zipf exponent equal to 1) to compare small sample biases in estimated β's across OLS, rank-adjusted OLS, and Theil estimators.[55] Figure B1 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 B1 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 Ibragi-

[54]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. Dietz (1989) considers a set of nonparametric slope estimators, and finds that the Theil estimator is robust, easy to compute, and competitive with alternative estimators in terms of mean squared error.

[55]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 B1: Monte Carlo Estimates of the Zipf Exponent
OLS, Rank−Adjusted OLS, and Theil Estimates



mov (2007) show that, when we estimate power law exponents in small samples, OLS standard errors are biased down.[56] The confidence interval associated with Theil regression estimates similarly overstates the estimator's precision when data are drawn from a distributional power law.[57] To gauge and compare the true precision of these estimators, we can use Monte Carlo simulations. **Figure B2** 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:

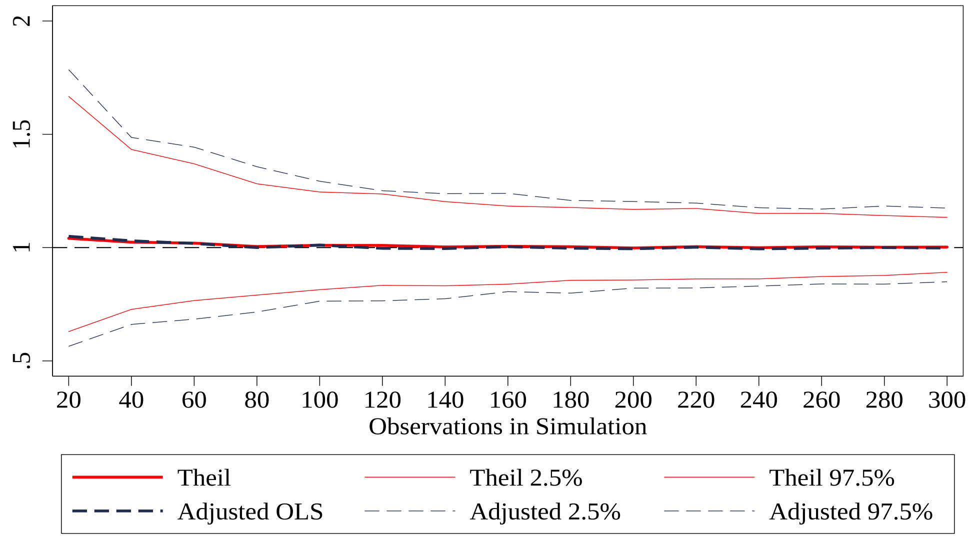
$$h \equiv (i,j) \quad X(h) \equiv \begin{bmatrix} 1 & x_i \\ 1 & x_j \end{bmatrix} \quad y(h) \equiv \begin{bmatrix} y_i \\ y_j \end{bmatrix} \quad b(h) \equiv X(h)^{-1} y(h)$$

Under this notation, the OLS estimator is: $\beta_{OLS} = \sum_{h=1}^{N} w(h)b(h)$, where the weights

[56] The true standard error of $\hat{\beta}$ in equation (22) is asymptotically $(2/n)^{0.5}\hat{\beta}$.

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

Figure B2: Monte Carlo Estimates of the Zipf Exponent
Comparison of Theil and Rank−Adjusted OLS
Mean Estimate and 95% Interval Over 1,000 Simulations



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."

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:

$$\ln R_i = \alpha - \beta \ln S_i + \epsilon_i \tag{23}$$

A number of studies suggest employing a regression augmented with a quadratic term to detect non-linearities and deviations from distributional power laws[58]:

$$\ln R_i = \beta_0 - \beta_1 \ln S_i + \beta_2 (\ln S_i)^2 + \nu_i \tag{24}$$

As discussed below, the standard errors associated with this model are biased down. However, I present historical estimates of equation (24) to facilitate comparison with existing studies using non-historical data. Table 4 shows that between 1500 and 1700,

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

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.

Table 4: 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. As discussed in the text, Table 5 corrects for the biases in these standard errors.

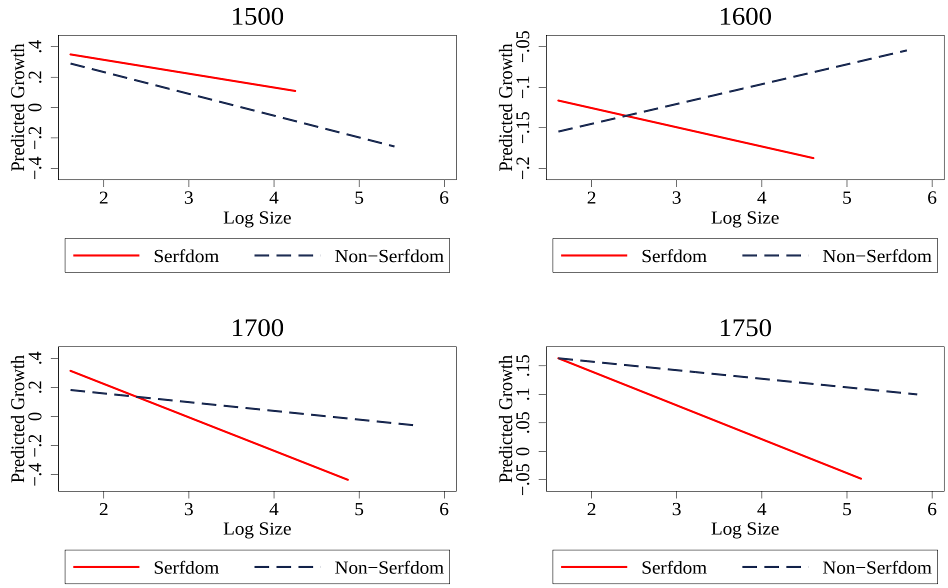
However, the estimates in Table 4 should be treated with caution. It can be shown using synthetic data from a pure power law distribution that heteroskedasticity-robust standard errors associated with equation (24) exhibit downward bias in finite samples.[59] It follows that the statistical significance of $\hat{\beta}_2$ is not a robust criterion on which to base rejection of Zipf's Law. Hence Table 4 should be read as indicating the existence (or absence) of gross departures from Zipf's Law, not as a precise test.

F Appendix: Serfdom

The consistency of the negative association between size and growth under serfdom is also evident when we examine the data on a period-by-period basis. To show this, I present

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

Figure 7: Predicted City Growth − Regression Analysis



Note: This figure shows the persistent and pronounced negative relationship between size and growth for cities exposed to the second serfdom. The regression model is: $\log S_{i,t+1} - \log S_{i,t} = \alpha_0 + \alpha_1 \log S_{i,t} + \alpha_2 CAPITAL_i + \epsilon_i$. The figure presents predicted values setting the capital indicator to zero.

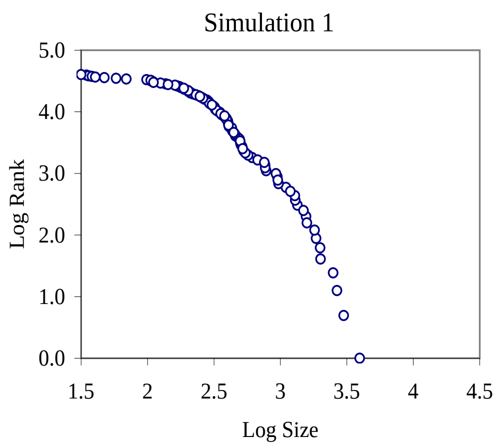
predictions from a regression where variations in the log of city growth are explained by the log of city size and political primacy (i.e. an indicator capturing whether or not a given city was a capital). **Figure 7** presents predicted values from this regression and shows that the negative correlations between city populations and city growth rates were pronounced and persistent in Eastern Europe over the second serfdom. However, the estimated relationship between size and growth under serfdom is not significant at conventional levels. This may be explained in two ways. First, it may be that the relationship is simply not statistically significant. Second, it may be that measurement error is attenuating the parameter estimates and masking a relatively strong underlying association. As suggested in Table 3 above, the data on city size is noisy. Since I calculate city growth based on observed size, observed city growth is likely very noisy.[60]

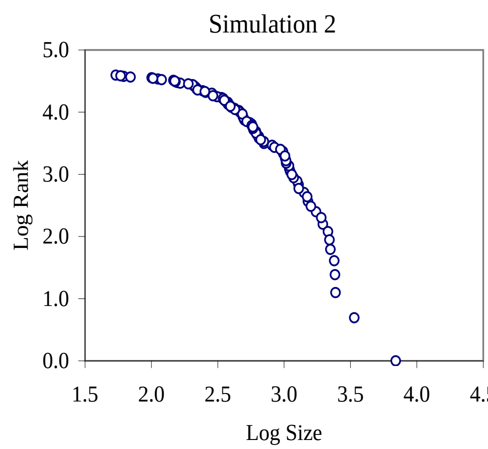
[60]Simple calibration exercises – not reported here – suggest that given unobserved data with a highly significant correlation growth and size, measurement error of the sort suggested by Table 3 would reduce the precision of the estimated association.

G Appendix: Simulation of Model

A simple – and provisional – simulation illustrates how the model can generate deviations from Zipf's Law. **Figure C1** shows the city-size distributions that result when one takes an arbitrary, fixed set of cities and runs them through the model assuming that the fixed land (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

Figure C1: City Sizes When Fixed Land Enters Production
Two Representative Simulations Based on City Growth Model





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.