Cities and Countries

Andrew K. Rose*

Draft: December 14, 2005
Preliminary: Comments Welcome.

Abstract

If one ranks cities by population, the rank of a city is inversely related to its size, a well-documented phenomenon known as Zipf's Law. Further, the growth rate of a city's population is uncorrelated with its size, another well-known characteristic known as Gibrat's Law. In this paper, I show that both characteristics are true of countries as well as cities; the size distributions of cities and countries are similar. But theories that explain the size-distribution of cities do not obviously apply in explaining the size-distribution of countries. The similarity of city- and country-size distributions is an interesting riddle.

Keywords: distribution; Zipf; Gibrat; law; empirical; mean; growth; rank; size; logarithm.

JEL Classification Numbers: FOO, R12

Contact: Andrew K. Rose, Haas School of Business,
University of California, Berkeley, CA 94720-1900
Tel: (510) 642-6609; Fax: (510) 642-4700
E-mail: arose@haas.berkeley.edu
URL: http://faculty.haas.berkeley.edu/arose

* B.T. Rocca Jr. Professor of International Business, Economic Analysis and Policy Group, Haas School of Business at the University of California, Berkeley, NBER Research Associate, and CEPR Research Fellow. I thank FRBNY and IMF for hospitality during the course of this research. For comments, I thank: Tony Braun, Jan Eeckhout, Jeff Frankel, Xavier Gabaix, Pierre-Olivier Gourinchas, Chad Jones, Volker Nitsch, Masao Ogaki, Eswar Prasad and workshop participants at the Universities of Pennsylvania, Tokyo and Washington. The data sets, key output, and a current version of the paper are available at my website.

1. Introduction

Cities are a standard unit of observation in urban economics, just as countries are a norm in international economics. The distribution of city sizes has been extensively studied. A couple of striking empirical regularities characterize the distribution of cities within a country. The rank (by size) of a city is almost perfectly inversely related to its size (at least for the largest cities), a stylized fact known as "Zipf's Law." It is also well known that growth in cities seems to be approximately proportionate, independent of city size; this is known as "Gibrat's Law." In this short paper, I consider both of these well-known characteristics of city size distributions, and show that they work about as well when one considers countries instead of cities.

2. Empirical Characteristics of City Size Distribution

The focus of this paper is a pair of well-known empirical regularities that characterize the distribution of population size across cities. "Zipf's Law" states that when cities are ranked by the size of their populations, city size is inversely correlated with rank. "Gibrat's Law" states that the size of a city is uncorrelated with its growth rate. Both stylized facts are long established, well known, and essentially undisputed to the best of my knowledge. Accordingly, I now briefly provide results that use recent data and are representative of the larger literature.

City Size and City Rank: Zipf's Law

"Zipf's law for cities" states that the number of cities with population greater than S is approximately proportional to 1/S.[1] The relationship fits well, and Zipf's Law characterizes the cities of different countries at different points of time.[2] A vast literature documents Zipf's Law, while a smaller literature attempts to explain it. Among the more recent references are Eeckhout

(2004), Gabaix (1999), Krugman (1996), and Rossi-Hansberg and Wright (2004); Gabaix and Ioannides (2004) provide a recent survey and Nitsch (2005) a recent meta-analysis.

Two methods have been used in the literature to document Zipf's Law: graphs and regressions. Both begin by ranking cities by the size of their population (New York is currently #1 in the United States, Los Angeles #2, and so forth). One then compares the natural logarithm of city rank to the natural logarithm of city population, using either a) graphical or b) regression techniques. Appendix Table A1 lists the populations of the largest American cities in 2000 (the most recent census).

Figure 1 presents a typical set of graphs. The top-left graph is a scatter-plot of the rank of the fifty largest America combined statistical areas (CSAs) in 2000 (on the ordinate or y-axis) against their sizes (on the abscissa or x-axis). Collectively the fifty CSAs covered almost 152 million people at the time of the 2000 census, around 56% of the population of the United States.[3] A line with slope of –1 is provided to facilitate comparison. Clearly Zipf's law works well. These results do not depend much on the exact year; the top-right graph is the analogue for 1990 census data. The exact definition of "city" does not matter much either; analogues for the 200 largest metropolitan and micropolitan statistical areas (MSAs) in 2000 and 1990 are provided in the bottom pair of graphs of Figure 1 (the 200 MSAs included almost 212 million people in 2000, over three-quarters of the population of the United States).[4]

Analogous regression results are tabulated in Table 1; these corroborate the graphical results. Each row in the table reports a regression of the log of city rank on the log of city size (and an unreported intercept).[5] Consider the results for the 50 largest CSAs in 2000, which are presented in the top row. The slope coefficient is -1.03, close to the Zipf value of –1.[6] I follow Gabaix and Ioannides (2004) and approximate the standard error by $\beta(2/N)^{.5}$ where β is the slope coefficient and N is the sample size; this delivers a standard error of .21 for the sample of 50 large cities.[7] The regression fits well, with an unadjusted R^2 of .98. Other lines in the table show that my results do not depend sensitively on the year, and that the largest MSAs give results similar to those of the largest CSAs.

A few other features are worthy of mention in passing. First, broadening the sample size to include more cities tends to lower the Zipf slope coefficient systematically, as is clear from the bottom part of Table 1. That is, there appear to be "too few" small cities for the entire distribution of cities to satisfy Zipf's law. This is a well-known tendency in the data; see e.g., Eeckhout (2004). Second, narrowing the definition of a city to that of a "Census Designated Place" (CDP) raises the Zipf coefficient, also consistent with Eeckhout (2004).[8] Finally these results tend to characterize cities outside the United States, consistent with Rosen and Resnick (1980). Zipf coefficient estimates are tabulated for 25 countries in appendix Table A2.[9]

City Size and City Growth: Gibrat's Law

"Gibrat's law for cities" says that the expected growth rate of cities is independent of city size. Gabaix (1999) has shown that there is a tight theoretical link between Gibrat's law and Zipf's law; if the former works well, the latter can be more easily understood. There is strong evidence in the literature that Gibrat's law is an empirical regularity that characterizes city growth; see, e.g., Eeckhout (2004) for recent evidence and references.

Gibrat's law works well for American cities. Graphical evidence is provided for American cities in Figure 2. Consider the top-left graph that scatters the 1990-2000 population growth rate of fifty largest (in 1990) American CSAs (on the ordinate) against their size in 1990 (on the abscissa).[10] There is no clear relationship between the (log) population of a city in 1990 and its (population) growth rate over the following decade. The non-parametric data smoother included in the diagram to "connect the dots" is essentially flat. The non-relationship between city size and city growth also characterizes all 113 CSAs (portrayed in the top-right), as well as the MSAs (in the bottom row).

Regression analysis confirms the visual impressions. In Table 2, I tabulate the slope coefficient (and associated robust standard error) from regressions of city population growth between 1990 and 2000 on the log of 1990 city population (and an unrecorded intercept). Different rows present results for the largest cities and all cities, using both CSA and MSA measures of city size. Three of the four slopes tabulated are insignificantly different from zero; there is a significant relationship between size and subsequent growth when the entire set of MSAs (but not CSAs) is included. All of the equations fit the data poorly.

3. Empirical Characteristics of Country Size Distribution

The preceding section summarized evidence that two stylized facts – Zipf's law and Gibrat's law – characterize the size distribution of cities. But the size distribution of countries has not been much studied, so far as I know. I now present results comparable to those of section 2, but for countries.

Country Size and Country Rank: Zipf's Law

Figure 3 is the analogue to Figure 1, but portrays large countries instead of large American cities. The figure presents scatter-plots of the (natural logarithm) of country rank against (log) country population. As with the cities, a line with slope of –1 is also provided.

There are nine graphics presented in the figure, portraying the size-rank distribution of the 50 largest countries at different years.

Just as there are different definitions of cities, there is no universal definition of a "country." Economists often ignore small countries such as Liechtenstein (just as they often ignore small cities like Clovis, NM).[11] Niue, a small island in the South Pacific, has been self-governing in free association with New Zealand since 1974 and had an estimated 2005 population of 2,166. Is it a country? As of July 2005, Tuvalu had a population estimated to be 11,636; Nauru had 13,048 people, and San Marino 28,880.[12] None of these countries has military forces, a currency, or an embassy in the United States, but all were members of the United Nations. Are any or all of these countries?

My default is to consider as "countries" all entities that are considered separately by standard data sources such as the 2005 *World Development Indicators* or the CIA *World Factbook*. This is easy and seems natural, since it adheres to the Ricardian notion that factors such as labor are more mobile within a country than between countries (just as labor is more mobile within a city than across cities). This also tends to coincide with the existence of a central government with a monopoly on legal coercion.[13] However, I do not rely on a single definition of "country." I also consider the set of independent sovereign nation states, so that I exclude entities considered as "countries" under my broad definition such as Hong Kong (a special administrative region of China), Puerto Rico (a commonwealth associated with the United States), Reunion (an overseas department of France), the Cayman Islands (a British crown colony) and the West Bank and Gaza strip (not internationally recognized as a *de jure* part of any country). Since most such entities are small, my results do not typically depend on the exact definition of "country."[14] The populations of the largest countries in 2004 are tabulated in Appendix Table A3.[15]

The 1900 data are taken from *The Statesman's Yearbook 1901* (a few missing observations are filled in from later editions); China was the largest country with 339.68 million inhabitants while the smallest of the 50 "countries" portrayed is Kamerun, a German colony with 3.5 million. In 1900 the largest 50 "countries" account for 1.407 billion people, over 92% of the world's population. The 1950 data and the 2050 projection are taken from the U.S. Census Bureau's *International Data Base*, while data for 1960 through 2000 are taken from the World Bank's *World Development Indicators*.[16] The 2004 populations were downloaded from the CIA's *World Factbook*.[17] In 2004, the largest 50 countries accounted for almost 5.6 billion people, around 88% of the world's population. Thus, the percentage of the total population covered is comparable to (indeed, slightly higher than) that of large American cities.

Figure 3 shows that Zipf's law works well for countries. The relationship between country rank and size is close to inverse and linear; the biggest exception is the cross-section from 1900. This is also true of independent sovereign nation states, portrayed in Figure A1.[18]

Corresponding regression results are tabulated in Table 3; I record the slope coefficients from a regression of (log) rank on (log) size (and an unrecorded intercept), along with the unadjusted R^2. None of the slopes are different from –1 at traditional confidence levels, and most are quite close to –1. The biggest exception is 1900, whose slope is slightly over one standard deviation away from -1. It is striking that the slopes are close to -1 in spite of the fact that the estimates are biased in small samples towards 0 (Gabaix and Ioannides, 2004). It is interesting to note that the slope coefficients rise over time. Most interesting of all is how similar the estimates are to the Zipf slopes for cities recorded in the top part of Table 1.

It is also interesting to compare two other empirical phenomena that have been documented for cities. First, as noted above and in Table 1, a broader sample of cities tends to be associated with a lower Zipf slope coefficient. The same is true of countries, as shown in table A4. This reports the number of "countries" (including dependencies, colonies, and so forth) with admissible population data, the Zipf slope coefficient and its standard error, and the unadjusted regression R^2. The slopes are substantially lower once all countries are included in the Zipf regression, and the fit deteriorates accordingly. Appendix figure A6 presents a scatter plot of country rank against country size in 2004, along with fitted lines for: 1) a Zipf regression estimated over the entire sample of (237) countries; 2) another Zipf regression estimated with only the 25 largest countries, and 3) a third regression fitted to the 150 largest countries.[19] The slope coefficient declines dramatically as smaller countries are included in the sample, consistent with Eeckhout (2004). Just as there are "too few" small cities, there are too few small countries.

On the other hand, the results are insensitive to the exact definition of "country." Appendix table A5 shows that these patterns also characterize the data when dependencies, territories, colonies, possessions, and so forth are excluded from the sample; it is the analogue to Table 3, but estimated only for independent sovereign nation states. 1900 is now even more of an exception, almost surely because colonies (like India, Indonesia, Nigeria, to name just the largest few) are treated as independent observations in table 3, but are excluded from Table A5.

Country Size and Country Growth: Gibrat's Law

Is the growth rate of a country's population tied closely to its size? Figure 4 examines this hypothesis; it is the analogue to Figure 2 for countries instead of cities. I take advantage of the fact that the *WDI* provides comparable data for a large number of "countries" starting in

1960. The top-left graph scatters growth between 1960 and 1970 against the log of the population in 1960. The other two graphs on the top row also start in 1960, but extend subsequent population growth to 1980 (in the middle) and 2000 (on the right). The middle row of graphs start in 1970, while the bottom row presents results that start in 1980 and 1990. Throughout, each country is marked by a single dot, and non-parametric data smoothers are included to "connect the dots." There are few signs of a strong consistent relationship between initial country size and subsequent population growth. The analogue for independent sovereign nation states is provided in Figure A3, while that for the 50 largest countries is in Figure A4.

To corroborate this impression more rigorously, regression analysis is provided in Table 4. Analogously to Table 2, I regress the population growth rate on the log of the initial population (and an unrecorded intercept); I report the slope coefficient, its robust standard error, and the R^2 of the regression. I vary the starting and ending years of the sample; I also use three different sets of countries (all countries, all sovereigns, and the 50 largest countries).

The effect of initial population on its subsequent growth is always estimated to be negative; smaller countries have faster population growth. That said, most of the slopes are insignificantly different from zero at standard confidence levels. The hypothesis of no effect of size on growth usually cannot be rejected, with exceptions when estimation starts in 1960, or when sovereigns are considered and estimations ends in 2000. None of the regressions fit well.

Country Size: Log Normality

Until recently, work on Zipf's law has focused on the largest cities. Eeckhout (2004) uses recent American census data that covers the populations of over 25,000 "Census Designated Places" in 2000. He finds that the distribution of places adheres closely to log-normality, as might be expected if Gibrat's Law works well. I now briefly investigate whether the distribution of "countries" is approximately log-normal.

Figure 5 provides histograms of the natural logarithm of country population at nine different years; the normal distribution is also portrayed to ease comparison. There is evidence of right-skewness; there are "too few" medium and small countries. Still, there is little evidence of kurtosis (fat tails). From an ocular viewpoint, log-normality fits reasonably well (Figure A5 is the analogue for sovereign nation states).

More rigorous examination of log-normality is provided in table 5. I tabulate p-values for the hypothesis of no excess skewness and kurtosis, both separately and jointly. Log-normality can be rejected for the first and last years I consider. On the other hand, it seems to be a reasonable description of the distribution of country populations from 1960 through 2000.

4. Empirical Regularity, Theoretical Puzzle

Suppose we believe that cities and countries have similar population distributions. What might explain this?

There has been little rigorous economic analysis of the size of countries. A notable exception is the body of work by Alesina and Spolaore summarized in their (2003) book. Alesina and Spolaore develop a theory of country size in which the benefits of size are offset by the costs of increased heterogeneity. Larger countries can supply public goods more inexpensively; they are also able to provide more regional insurance and income redistribution. Productivity may also be higher in larger countries because specialization is limited by the size of the market (especially for closed economies). But Alesina and Spolaore point out that these benefits may be costly, since larger countries also tend to be more heterogeneous. But the focus of Alesina and Spolaore is on the size of a representative country; they do not directly study the size distribution of countries *per se*.

In contrast, there has been much professional interest in the determination and distribution of city sizes; see, e.g., Eeckhout (2004), Gabaix (1999), Krugman (1996), and Rossi-Hansberg and Wright (2004). Theories of city size typically balance the positive effects of agglomeration against negative externalities. The former can result from e.g., knowledge spillovers or scale economies, while the latter can arise from congestion, commuting, or land prices. Both types of externalities are required to induce mobile labor to migrate between cities in appropriate proportions. Krugman (1996) surveys a number of different theories that rationalize Zipf's Law for cities and finds (p 401) "it impossible to be comfortable with the present state of our understanding."

In Eeckhout (2004), the positive local production spillovers are offset by congestion and higher property prices. Cities receive exogenous technology shocks, and identical workers are free to choose between cities with high productivity, wages, property prices, and commuting times and cities that have lower values for these variables. Rossi-Hansberg and Wright (2004) is related but uses externalities and shocks at the industry level. Gabaix (1999) stresses the role of shocks to a city's amenities in inducing migration between cities; these may be man-made (e.g., shocks to the environment, judicial system, or transportation network) or natural (e.g., natural disasters or weather). With independent and identically distributed amenity shocks, both Gibrat's and Zipf's Law are satisfied.

None of this work seems readily extendible to countries. Cities and countries are different phenomena in a number of different aspects. For one, countries have more control over their policies and institutions than cities. Since many features of life and work are determined at the national level, cities within a country are more similar than different countries. Mobility is much higher between cities inside a country than it is between countries, so that theories in which workers choose their city of residence seem inappropriate to countries.[20] Externalities, agglomeration effects, and amenity shocks that seem reasonable at a local level are less plausible at the national level. It is challenging to use theories of city dispersion to explain country sizes.[21]

On the other hand, since the size distribution of cities and countries are similar, it is natural to imagine that the same theory might explain both. One is left with the feeling that some deeper theory is required to explain this empirical regularity. If my empirical findings are corroborated, they constitute an intriguing puzzle for future theoretical work.

5. Conclusion

In this short paper, I have investigated the size distribution of countries' population. I have shown that it adheres reasonably well to Zipf's law (size and rank are inversely linked), and Gibrat's law (population growth rates are uncorrelated with size). These features, and other phenomena, are akin to more familiar characteristics of the size distribution of cities. Indeed, this paper is easily summarized in Figure 6, which compares city and country features directly.

Cities and countries have similar size distributions. This resemblance naturally suggests that a common explanation explains the size distribution of both cities and countries. But the only theoretical work in this area has focused on rationalizing the size distribution of cities, and models of the size distribution of cities do not seem easily applicable to countries. The common empirical regularities of cities and countries pose an interesting riddle for economics.

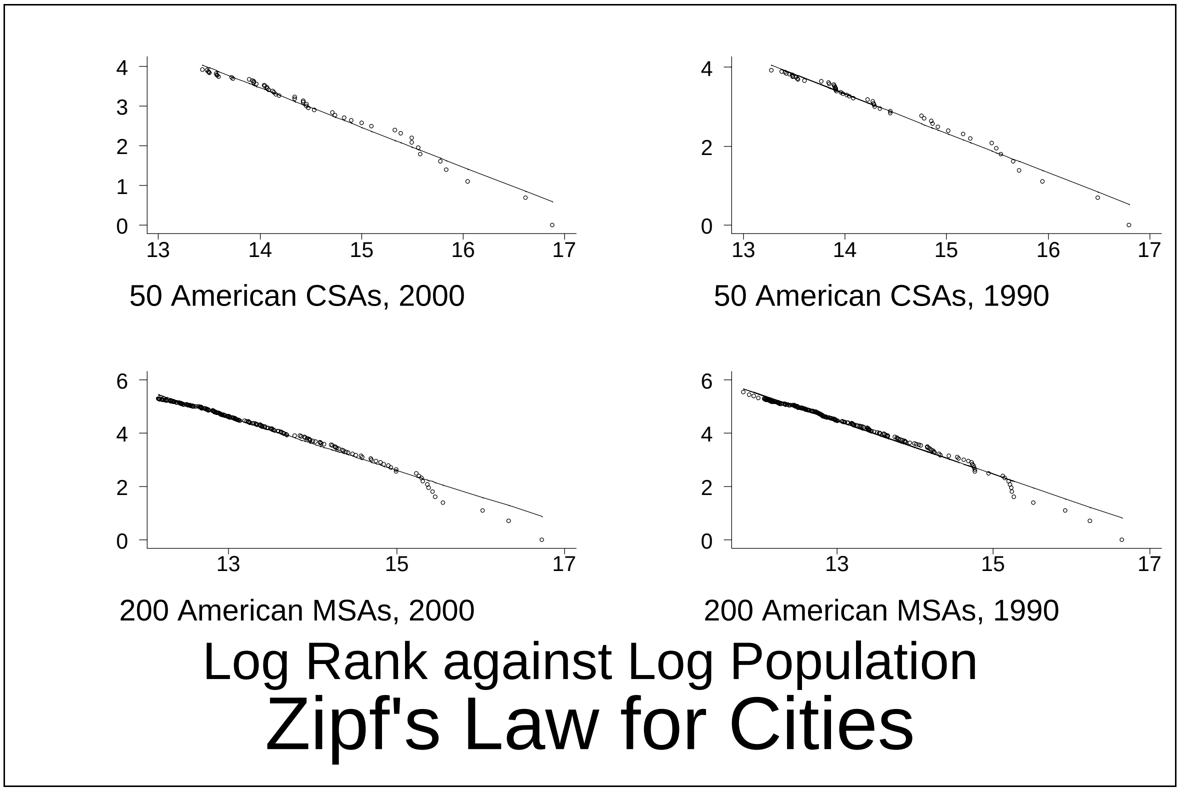

Figure 1: Size Distribution of Cities

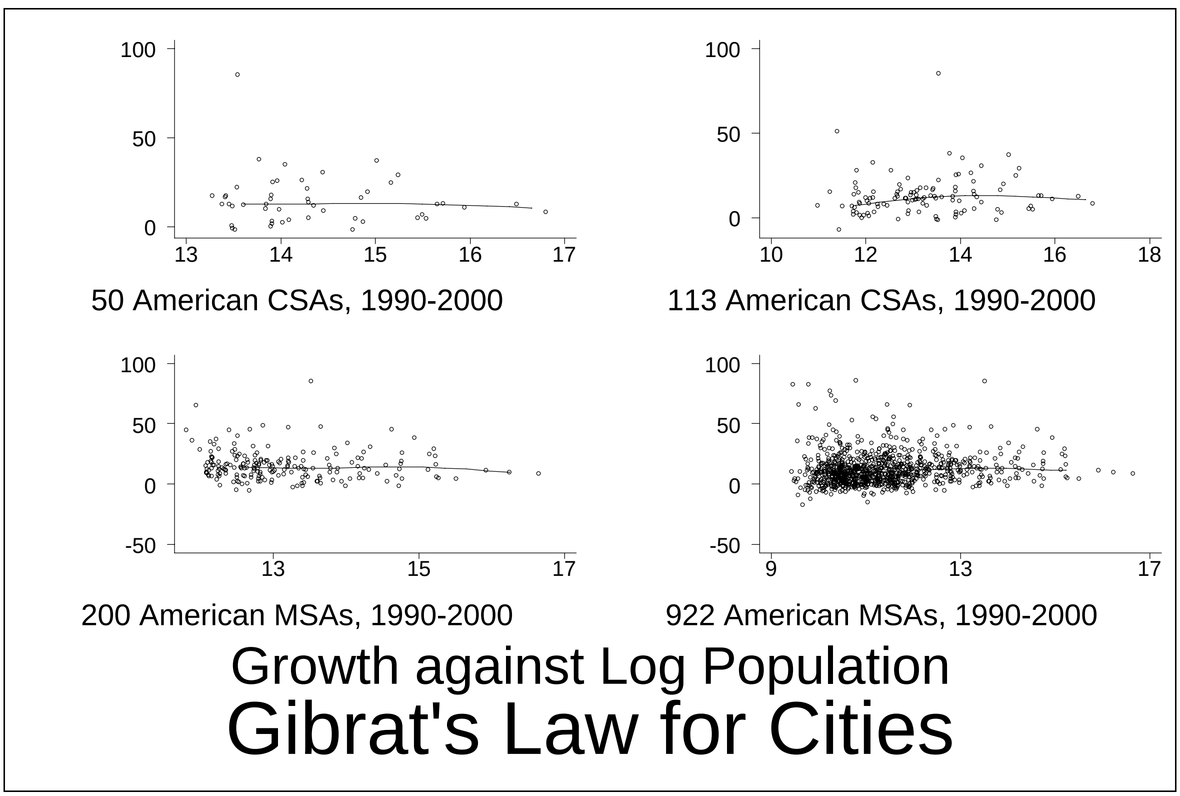

Figure 2: City Population Growth Rates

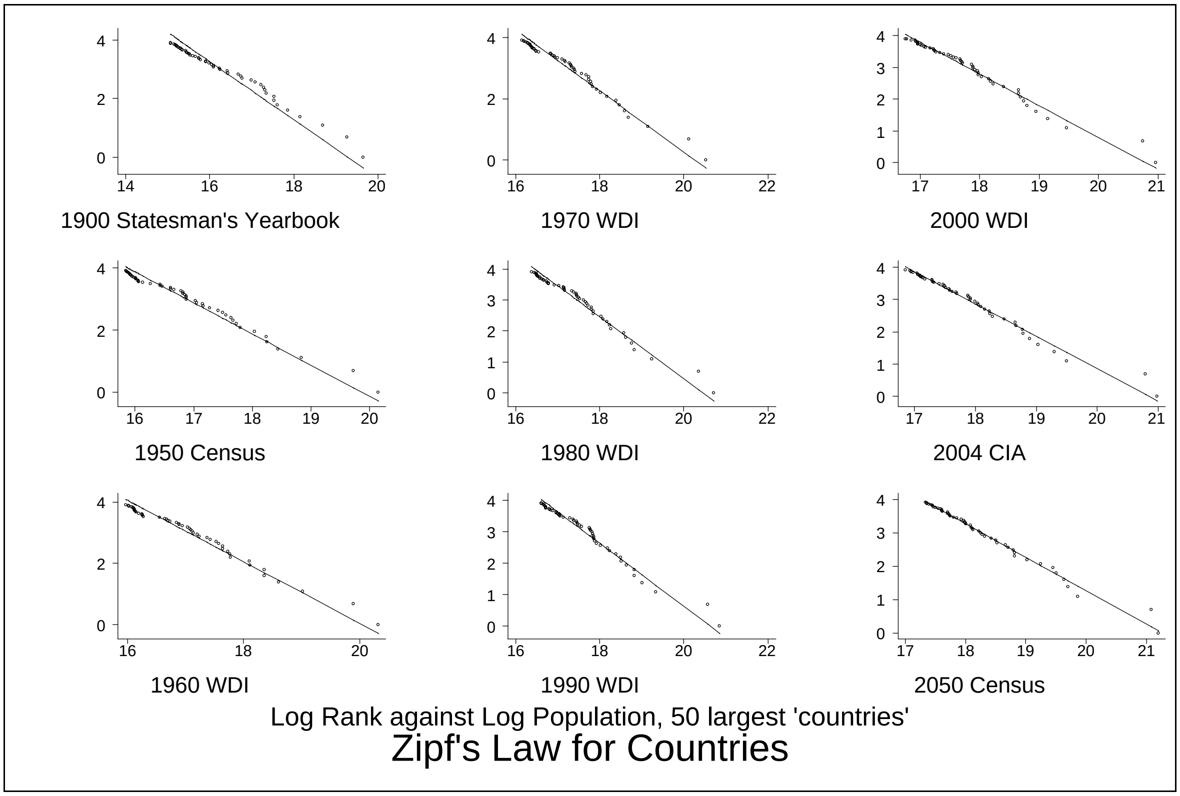

Figure 3: Size Distribution of Countries

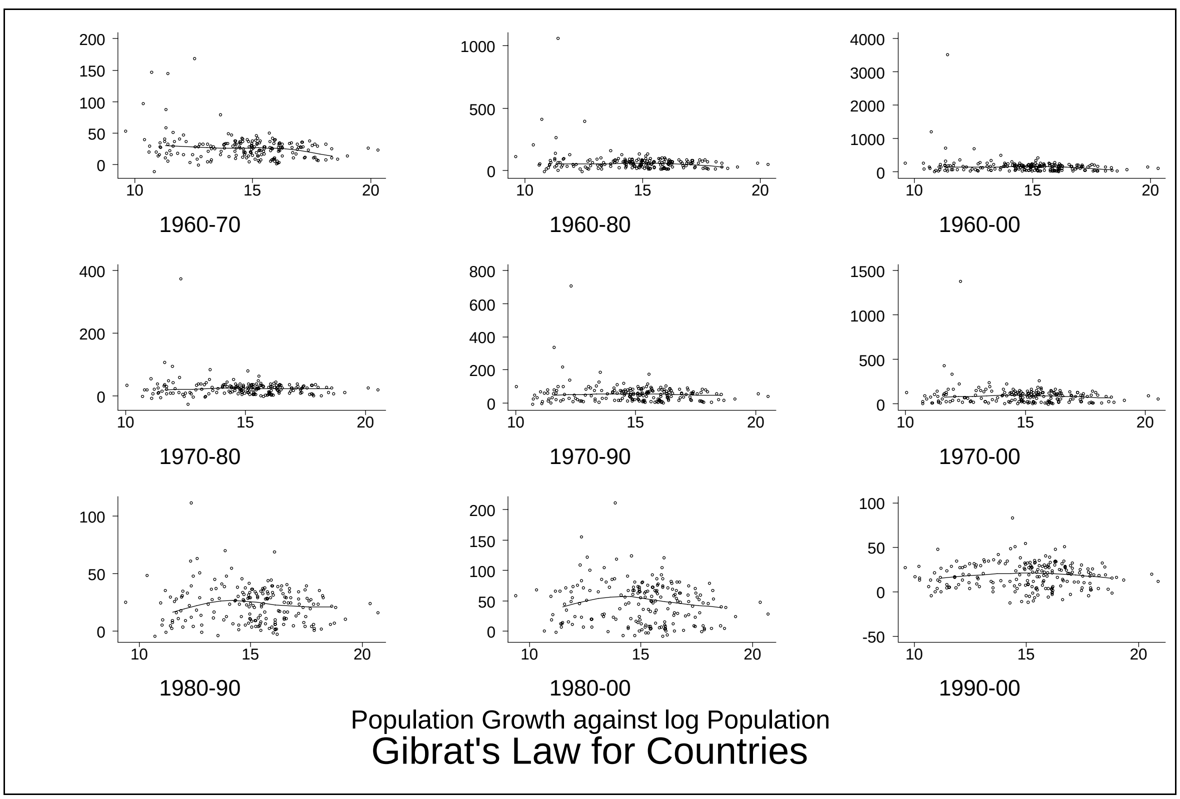

Figure 4: Country Population Growth Rates

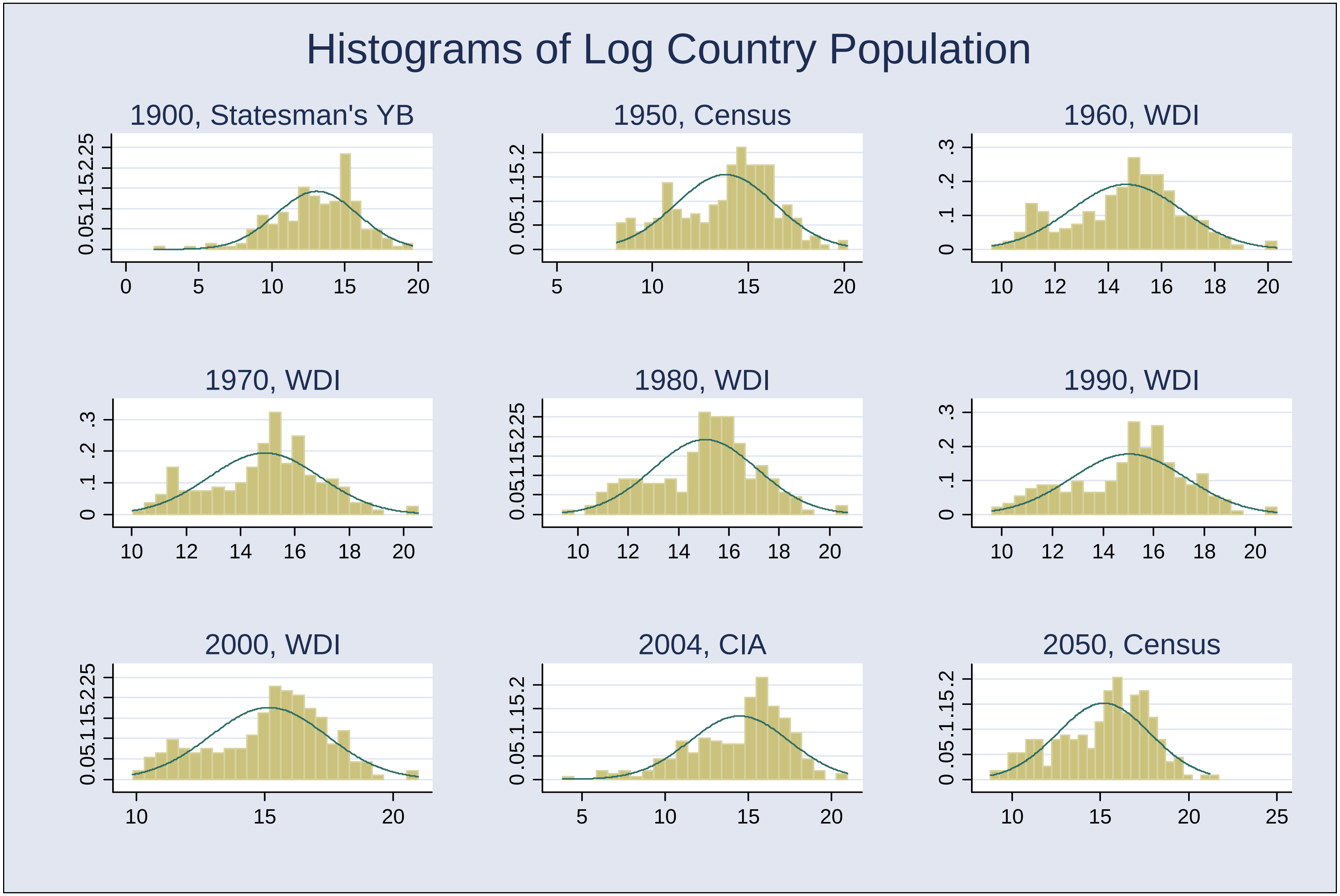

Figure 5: Histograms of the Natural Logarithm of Country Population

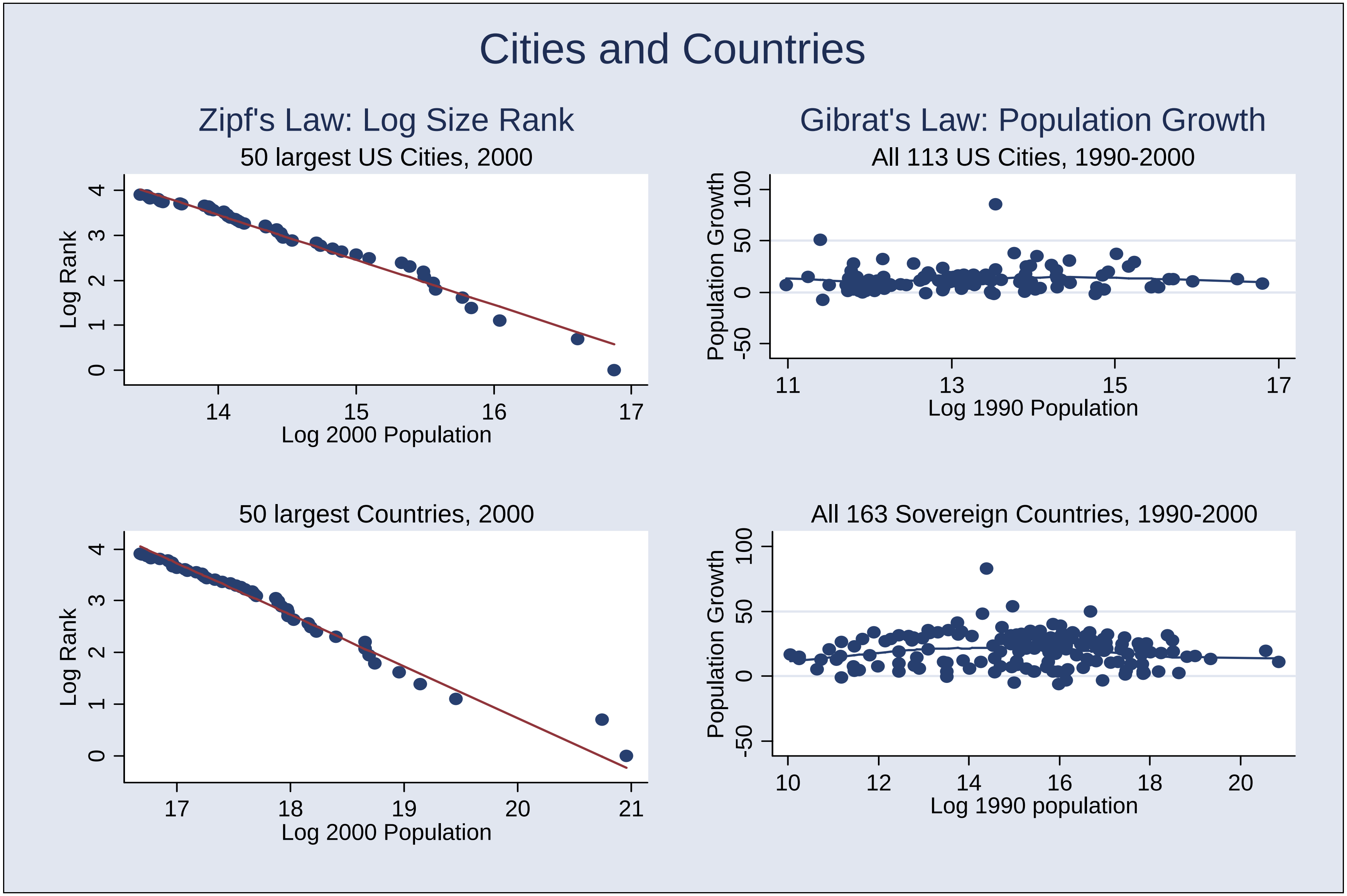

Figure 6: Summary of the Size Distribution of Cities and Countries

Table 1: Zipf Coefficients for Large American Cities

Year	City Measure	Sample	Slope (se)	R^2
2000	CSAs	50	-1.03 (.21)	.98
1990	CSAs	50	-1.03 (.21)	.98
2000	MSAs	200	-1.01 (.1)	.98
1990	MSAs	200	-1.02 (.1)	.98
2000	CSAs	113	-.73 (.10)	.93
1990	CSAs	113	-.74 (.10)	.93
2000	MSAs	922	-.82 (.04)	.98
1990	MSAs	922	-.83 (.04)	.98
2000	CDPs	601	-1.34 (.08)	.998

Coefficients are slopes from OLS regressions of log rank on log population.
Intercepts included but not recorded.
Approximate standard errors $(=\beta\sqrt{2/N})$.
"CSAs" denotes "Combined Statistical Areas" "MSAs" denotes "Metropolitan and Micropolitan Statistical Areas and "CDPs" denotes "Census Designated Places"

Table 2: Gibrat Coefficients for Large American Cities

City Measure	Sample	Slope (se)	R^2
CSAs	50	-1.48 (2.08)	.01
MSAs	200	-.01 (.78)	.00
CSAs	113	.97 (.82)	.01
MSAs	922	1.07** (.39)	.01

Coefficients are slopes from OLS regressions of population growth between 1990-2000 on log 1990 population.
Intercepts included but not recorded. Robust standard errors in parentheses.
* (**) indicates that the coefficient is significantly different from zero at the .05 (.01) level.
"CSAs" denotes "Combined Statistical Areas" "MSAs" denotes "Metropolitan and Micropolitan Statistical Areas and "CDPs" denotes "Census Designated Places"

Table 3: Zipf Coefficients for 50 Largest Countries

Year	Slope (se)	R^2
1900	-.78 (.16)	.99
1950	-.87 (.17)	.99
1960	-.88 (.18)	.98
1970	-.89 (.18)	.98
1980	-.91 (.18)	.98
1990	-.93 (.19)	.98
2000	-.95 (.19)	.98
2004	-.96 (.19)	.98
2050	-.99 (.20)	.99

Coefficients are slopes from OLS regressions of log rank on log population.
Intercepts included but not recorded.
Approximate standard errors $(=\beta\sqrt{2/N})$.

Table 4: Gibrat Coefficients for Countries

		All Countries	All Countries	All Sovereigns	All Sovereigns	Top 50	Top 50
Initial Year	**Final Year**	**Slope (se)**	**R^2**	**Slope (se)**	**R^2**	**Slope (se)**	**R^2**
1960	1970	-2.8** (1.0)	.08	-.7 (.8)	.01	-1.1 (1.4)	.01
1960	1980	-9.25* (4.6)	.05	-5.0* (2.0)	.02	-3.0 (3.3)	.01
1960	1990	-17.2* (8.4)	.05	-9.8** (3.5)	.07	-5.1 (5.7)	.01
1960	2000	-26.6 (14.4)	.04	-20.3** (5.6)	.11	-8.9 (8.5)	.02
1970	1980	-1.8 (1.36)	.01	-1.3 (.9)	.02	-2.7 (1.5)	.05
1970	1990	-4.3 (2.7)	.02	-2.9 (1.8)	.02	-6.1 (3.5)	.05
1970	2000	-7.8 (4.9)	.02	-7.3* (3.0)	.02	-11.2 (5.8)	.06
1980	1990	-.8 (.6)	.01	-.9 (.7)	.01	-2.4 (1.4)	.04
1980	2000	-1.7 (1.1)	.01	-3.6* (1.5)	.04	-6.1 (3.2)	.05
1990	2000	-.1 (.4)	.00	-1.2 (.7)	.02	-2.5 (1.6)	.04

Coefficients are slopes from OLS regressions of population growth (initial to final year) on log initial population. Intercepts included but not recorded. Robust standard errors in parentheses.
* (**) indicates that the coefficient is significantly different from zero at the .05 (.01) level.

Table 5: Tests for Normality in Country Log(Population) Distribution

	All Skewness	All Kurtosis	All Joint Test	Sovereigns Skewness	Sovereigns Kurtosis	Sovereigns Joint Test
1900, SYB	.00	.04	.00	.23	.25	.23
1950, Census	.05	.01	.01	.00	.01	.00
1960, WDI	.23	.39	.33	.13	.29	.17
1970, WDI	.26	.42	.38	.73	.41	.67
1980, WDI	.14	.66	.30	.38	.66	.61
1990, WDI	.05	.27	.08	.07	.57	.16
2000, WDI	.05	.17	.06	.02	.73	.05
2004, CIA	.00	.14	.00	.00	.03	.00
2050, Census	.01	.03	.01	.00	.74	.01

P-values shown are tests for hypothesis of no excess skewness and/or kurtosis; low values are inconsistent with hypothesis of log-normality.

References

Alesina, Alberto and Enrico Spolaore (2003) *The Size of Nations* (MIT Press).

Axtell, Robert L. (2001) “Zipf Distributions of U.S. Firm Sizes” *Science* 293, 1818-1820.

Eeckhout, Jan (2004) “Gibrat’s Law for (All) Cities” *American Economic Review* 94-5, 1429-1451.

Gabaix, Xavier (1999) “Zipf’s Law for Cities” *Quarterly Journal of Economics* 114-3, 739-767.

Gabaix, Xavier and Yannis M. Ioannides (2004) “The Evolution of City Size Distributions” in *The Handbook of Regional and Urban Economics* 4 (North-Holland; Henderson and Thisse, eds), 2341-2378.

Krugman, Paul (1996) “Confronting the Mystery of Urban Hierarchy” *Journal of the Japanese and International Economies* 10, 399-418.

Nitsch, Volker (2005) “Zipf Zipped” *Journal of Urban Economics* 57, 86-100.

Rosen, Kenneth T. and Mitchel Resnick (1980) “The Size Distribution of Cities” *Journal of Urban Economics* 8, 165-186.

Rossi-Hansberg, Esteban and Mark Wright (2004) “Urban Structure and Growth” unpublished.

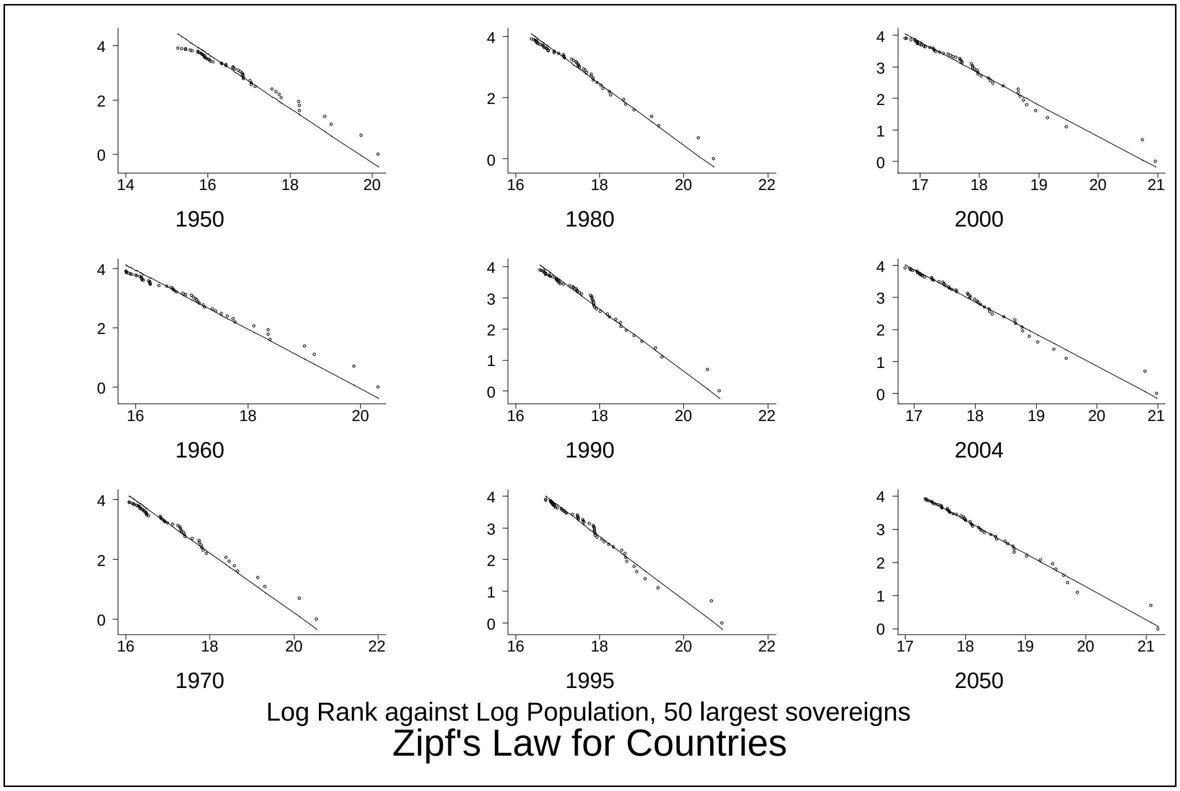

Figure A1: Size Distribution of Independent Sovereign Nation States

Figure A2: Size Distribution of Density of Countries

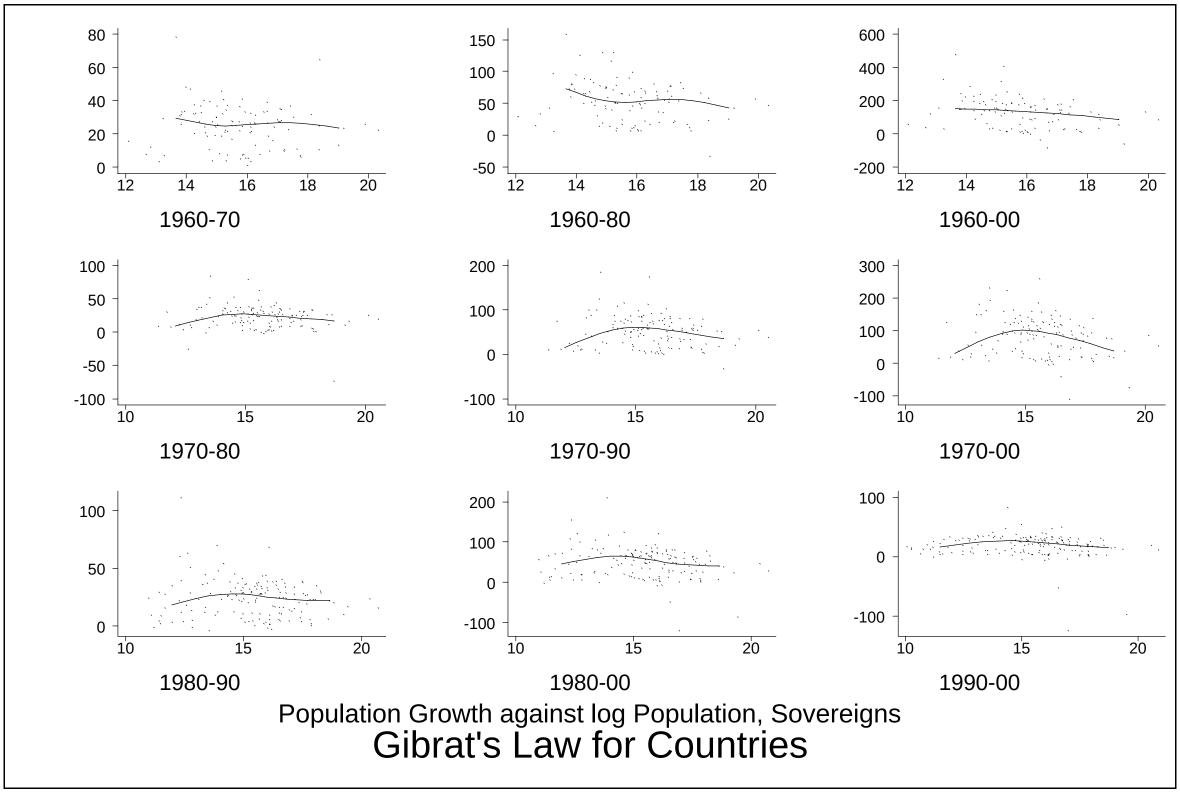

Figure A3: Country Population Growth Rates

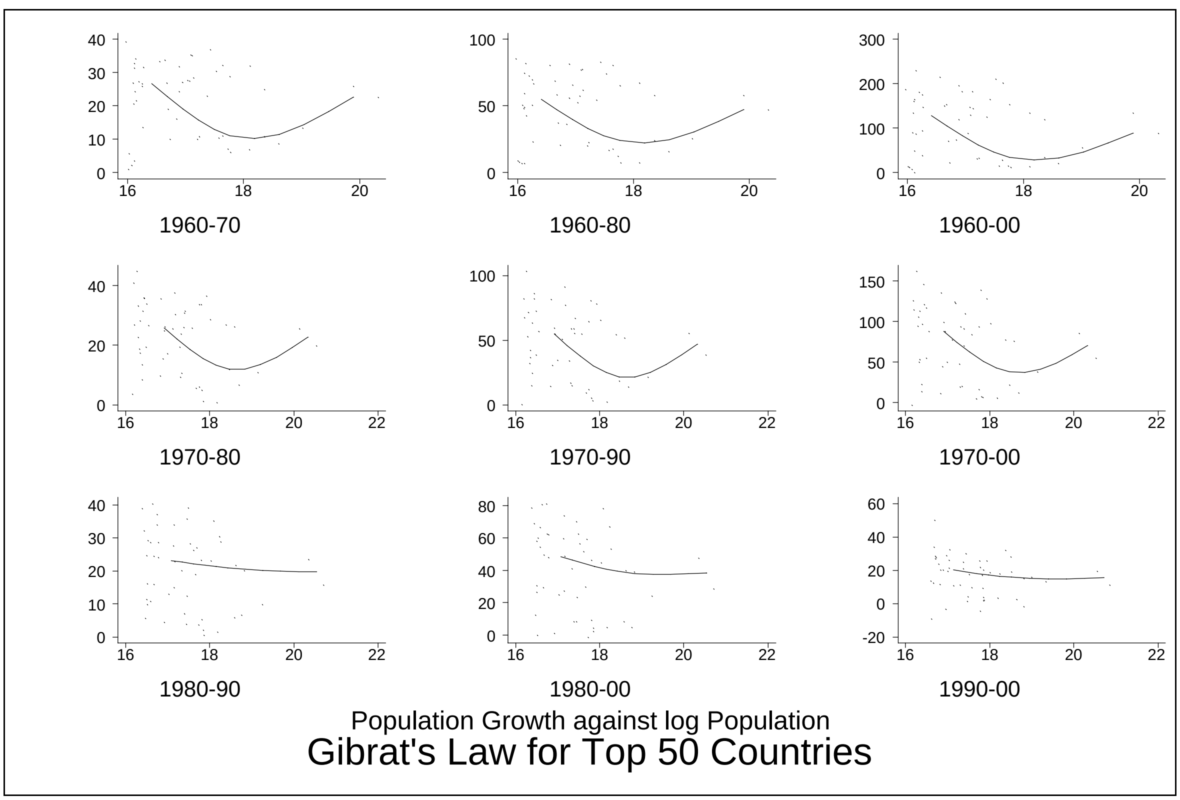

Figure A4: Country Population Growth Rates

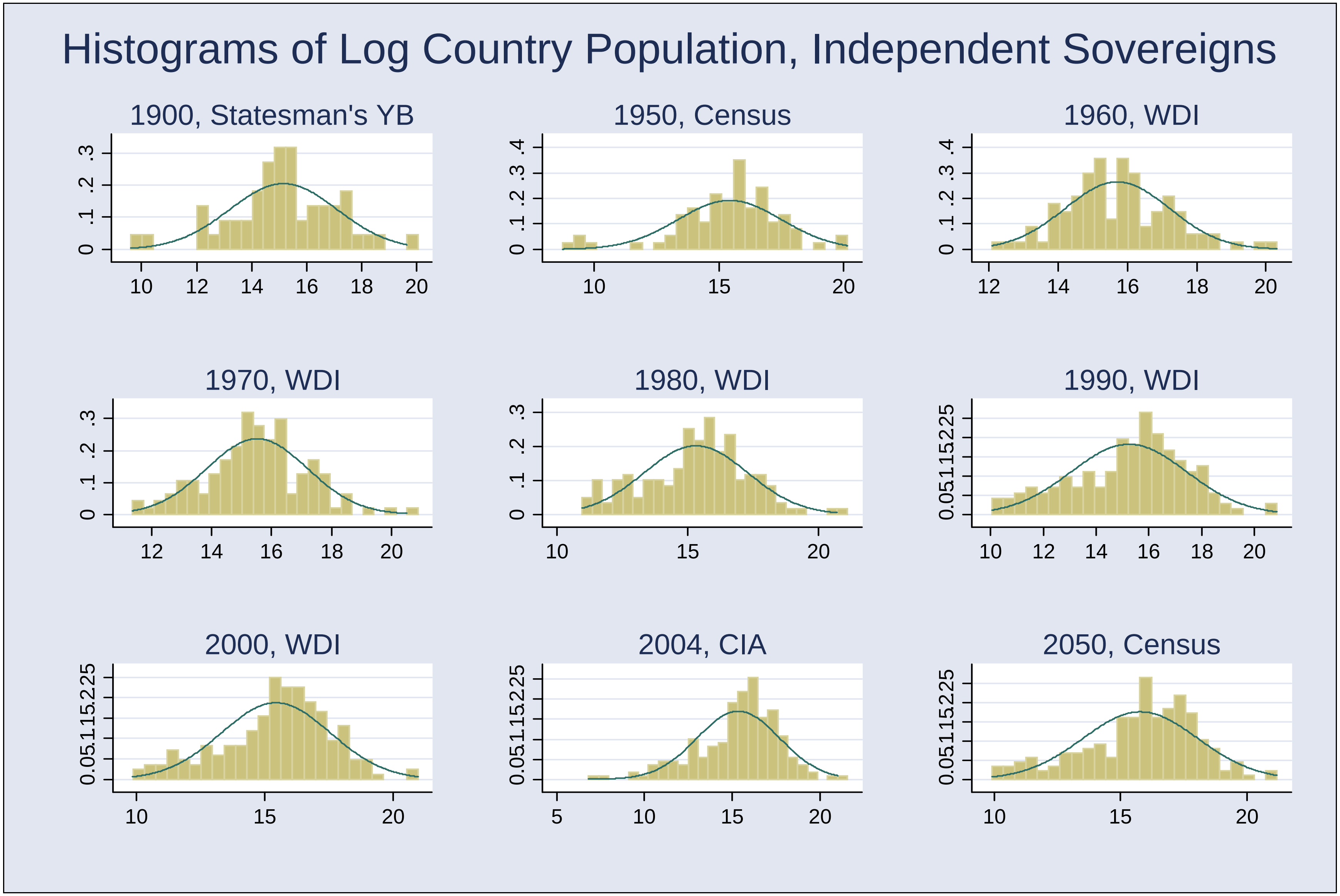

Figure A5: Histograms of Log Population Independent Sovereign Nation State

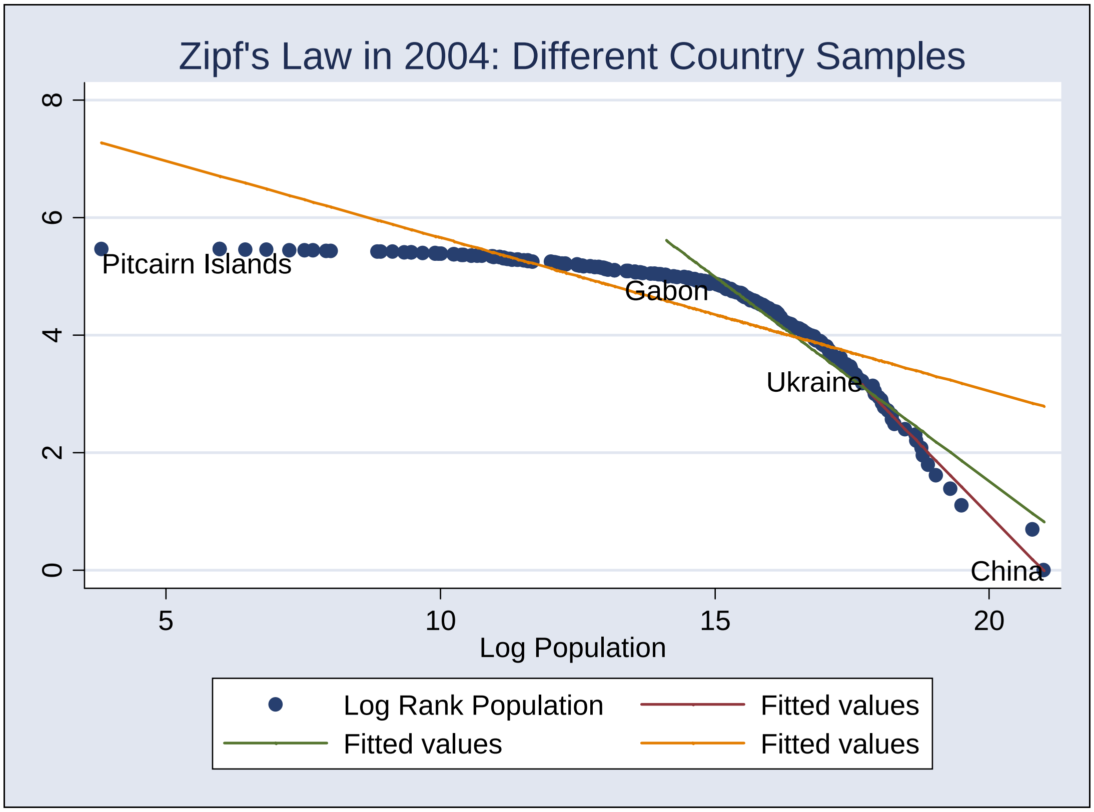

Figure A6: Country Size-Ranks with Varying Cutoff Points, 2004

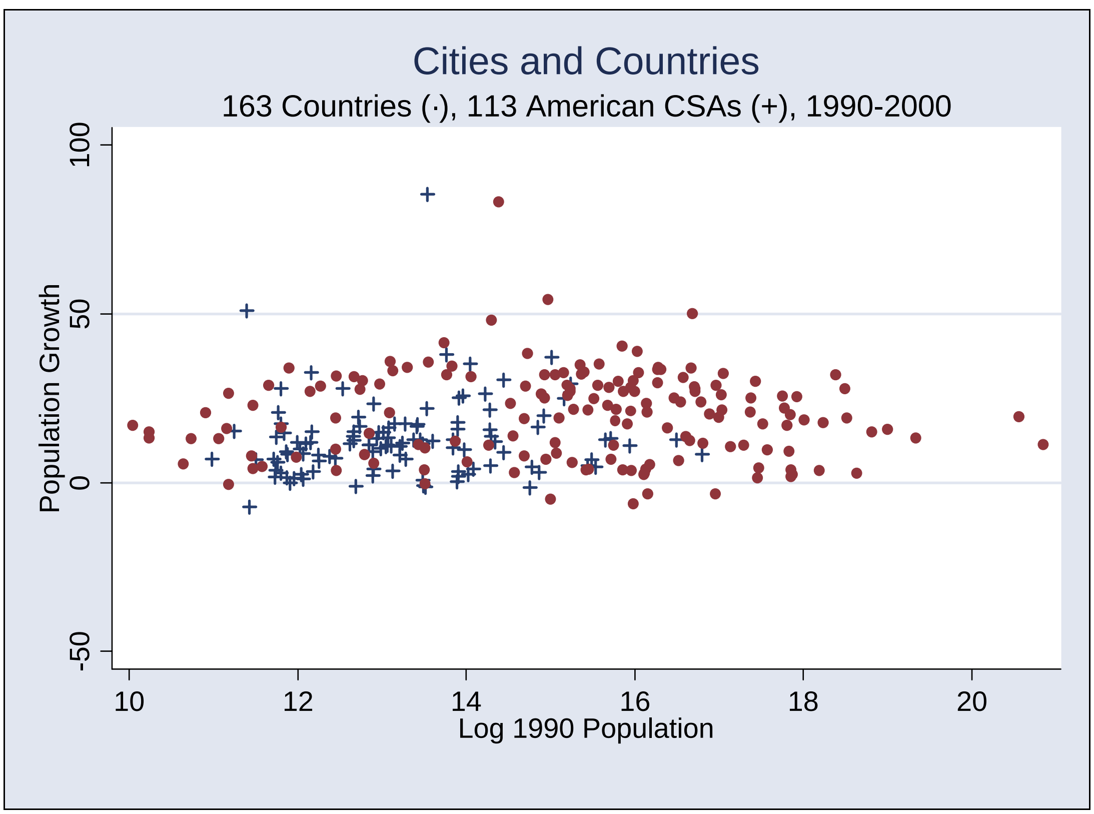

Figure A7: City and Country Population Growth Rates

Table A1: Large American City Populations, 2000

	Combined Statistical Area	Metropolitan Statistical Area	Census Designated Place
New York	21,361,797	18,323,002	8,008,278
Los Angeles	16,373,645	12,365,627	3,694,820
Chicago	9,312,255	9098,316	2,896,016
Washington	7,538,385	4,796,183	572,059
San Francisco	7,092,596	4,123,740	776,733
Philadelphia	5,833,585	5,687,147	1,517,550
Boston	5,715,698	4,391,344	589,141
Detroit	5,357,538	4,452,557	951,270
Dallas	5,346,119	5,161,544	1,188,580
Houston	4,815,122	4,715,407	1,953,631
Atlanta	4,548,344	4,247,981	416,474
Seattle	3,604,165	3,043,878	563,374
Minneapolis	3,271,888	2,968,806	382,618
Cleveland	2,945,831	2,148,143	478,403
St. Louis	2,754,328	2,698,687	348,189
Pittsburgh	2,525,730	2,431,087	334,563
Denver	2,449,054	2,179,240	554,636
Cincinnati	2,050,175	2,009,632	331,285
Sacramento	1,930,149	1,796,857	407,018
Kansas City	1,901,070	1,836,038	441,545
Charlotte	1,897,034	1,330,448	540,828
Indianapolis	1,843,588	1,525,104	781,870
Columbus	1,835,189	1,612,694	711,470
Orlando	1,697,906	1,644,561	185,951
Milwaukee	1,689,572	1,500,741	596,974
Salt Lake City	1,454,259	968,858	181,743
Las Vegas	1,408,250	1,375,765	478,434
Nashville	1,381,287	1,311,789	545,524
New Orleans	1,360,436	1,316,510	484,674
Raleigh	1,314,589	797,071	276,093
Louisville	1,292,482	1,161,975	256,231
Greensboro	1,283,856	643,430	223,891
Hartford	1,257,709	1,148,618	121,578
Grand Rapids	1,254,661	740,482	197,800
Oklahoma City	1,160,942	1,095,421	506,132
Rochester	1,131,543	1,037,831	219,773
Birmingham	1,129,721	1,052,238	242,820
Albany	1,118,095	825,875	95,658
Dayton	1,085,094	848,153	166,179
Fresno	922,516	799,407	427,652

Table A2: Zipf Slope Coefficients for Large Cities in Different Countries

Country	Top 50 Cities	All Cities	Top 50 Urban Agglomerations (se≈.20)	All Urban Agglomerations
Argentina		-1.07 (.26)		
Brazil	-1.23 (.25)	-1.23 (.11)		
Canada		-1.05 (.23)		
China	-1.49 (.30)	-1.34 (.08)		
Colombia		-.99 (.21)		
France		-1.36 (.33)		
Germany	-1.36 (.27)	-1.28 (.20)		
India	-1.31 (.26)	-1.17 (.08)	-1.08 (.22)	-.93 (.08)
Indonesia		-.90 (.18)		
Iran		-1.03 (.20)		
Italy		-1.15 (.25)		
Japan	-1.29 (.26)	-1.31 (.12)		
Korea		-.91 (.17)		
Mexico		-1.02 (.23)		
Nigeria		-1.06 (.19)		
Pakistan		-.86 (.17)		
Philippines		-1.13 (.24)		
Poland		-1.31 (.29)		
Russia	-1.42 (.28)	-1.18 (.13)	-1.32 (.26)	-1.15 (.18)
Spain		-1.36 (.26)		
Thailand				-1.18 (.24)
Turkey		-1.04 (.22)		
UK	-1.48 (.30)	-1.94 (.17)		
USA	-1.37 (.27)	-1.33 (.12)	-1.17 (.23)	-.85 (.11)
Ukraine		-1.14 (.24)		

Coefficients are slopes from OLS regressions of log rank on log population.
Intercepts included but not recorded.

Approximate standard errors $(= \beta\sqrt{2/N})$.

Table A3: Large Country Populations, 2004

China	1,298,847,616
India	1,065,070,592
United States	293,027,584
Indonesia	238,452,960
Brazil	184,101,104
Pakistan	159,196,336
Russia	143,974,064
Bangladesh	141,340,480
Japan	127,333,000
Nigeria	125,750,352
Mexico	104,959,592
Philippines	86,241,696
Vietnam	82,662,800
Germany	82,424,608
Egypt	76,117,424
Ethiopia	71,336,568
Turkey	68,893,920
Iran	67,503,208
Thailand	64,865,524
France	60,424,212
United Kingdom	60,270,708
Congo (Kinshasa/Zaire)	58,317,928
Italy	58,057,476
Korea, South	48,233,760
Ukraine	47,732,080
South Africa	44,448,472
Burma/Myanmar	42,720,196
Colombia	42,310,776
Spain	40,280,780
Sudan	39,148,160
Argentina	39,144,752
Poland	38,626,348
Tanzania	36,070,800
Kenya	32,982,108
Canada	32,507,874
Morocco	32,209,100
Algeria	32,129,324
Afghanistan	28,513,676
Peru	27,544,304
Nepal	27,070,666

Table A4: Zipf Coefficients for All Countries

Year	Number of Countries	Slope (se)	R^2
1900	205	-.29 (.03)	.73
1950	227	-.32 (.03)	.75
1960	191	-.40 (.04)	.78
1970	192	-.41 (.04)	.79
1980	194	-.40 (.04)	.78
1990	205	-.37 (.04)	.75
2000	207	-.37 (.04)	.75
2004	237	-.26 (.02)	.66
2050	227	-.31 (.03)	.72

Coefficients are slopes from OLS regressions of log rank on log population.
Intercepts included but not recorded.
Approximate standard error $(= \beta\sqrt{2/N})$.

Table A5: Zipf Coefficients for Independent Sovereign Countries

Sample	Slope (se)
1900	-.55 (.11)
1950	-.79 (.16)
1960	-.83 (.17)
1970	-.86 (.17)
1980	-.90 (.18)
1990	-.91 (.18)
1995	-.95 (.19)
2000	-.95 (.19)
2004	-.96 (.19)
2050	-.99 (.20)

Coefficients are slopes from OLS regressions of log rank on log population.
50 Countries included; intercepts included but not recorded.
Approximate standard error=.2 $(= \beta\sqrt{2/N})$.

Endnotes

[1] More rigorously, if one ranks large cities by population size, $S_1>S_2>\ldots S_N$, then $P(Size>S)\approx\alpha S^{-\beta}$ where α is a constant and $\beta\approx1$.
[2] Zipf's law also works well for firms; see Axtell, 2001.
[3] The smallest CSA portrayed is Columbia-Newberry South Carolina with a population of 519,415. The United States had 113 CSAs in 2000, the smallest being Clovis New Mexico with population 63,062. Data and details are available at http://www.census.gov/population/cen2000/phc-t29/tab06.xls
[4] The smallest MSA portrayed is Tuscaloose Alabama with population 192,034 in 2000. In 2000, the United States had 922 MSAs, the smallest being Andrews Texas with population 13,004. Data and details are available at http://www.census.gov/population/cen2000/phc-t29/tab03a.xls
[5] Using the notation of note 1, I use OLS to estimate $\ln(i)=\alpha+\beta\ln(S_i)+\varepsilon_i$ where ε is a disturbance term hopefully orthogonal to ln(S); Table 1 presents estimates of β.
[6] This is especially true when the negative bias documented by Gabaix and Ioannides (2004) is taken into account.
[7] The more conventional robust OLS standard error for the slope is .04.
[8] CDPs are only available for the year 2000. Data and details are available at http://www.demographia.com/db-uscity98.htm
[9] These were generated with UN data from http://unstats.un.org/unsd/demographic/products/dyb/DYB2002/Table08.xls. Combining together cities from different countries also delivers a Zipf slope coefficient of around -1, with or without country-specific fixed effects. However, countries may measure cities in different ways, so pooling data for a joint Zipf regression is problematic.

I note in passing that it is hard to pool data in a meaningful way for a Zipf regression across cities and countries because large countries are bigger than large cities, so the top end of the distribution is dominated by countries. However, pooling together data from the American cities and sovereign countries depicted in Figure 6 delivers the Gibrat graph in Figure A7.
[10] Nothing changes if the fifty largest cities in 2000 are considered.
[11] Visitors are welcomed to Liechtenstein and Clovis respectively at http://www.liechtenstein.li/ and http://www.cityofclovis.org/.
[12] http://www.cia.gov/cia/publications/factbook/
[13] The correspondence is imperfect for both criteria. Labor is mobile between countries over long periods of time. Dependencies do not have a complete monopoly over legal coercion (though since mother countries rarely exercise their rights, it is typically a *de facto* near-complete monopoly); neither do sovereign nation states (think of the Korean or first Gulf wars).
[14] Of course, the number of countries varies over time; Alesina and Spolaore (2003) provide more analysis.
[15] Another issue is that countries (like cities) change physical size over time. Czechoslovakia, Ethiopia, Pakistan, the USSR and Yugoslavia have split into multiple countries (and East Timor has split from Indonesia); the Cameroons, Germanies, Yemens have merged (as has Tanzania). My WDI data use countries defined as of 2005, so that e.g., they merge East and West Germany population for the period before unification in 1990. Dropping such countries that have merged/split has little impact on the Zipf coefficients reported in Table 3 below.
[16] The former is available at http://www.census.gov/ipc/www/idbrank.html while the latter can be obtained at http://www.worldbank.org/data/wdi2005/
[17] The CIA data set is available at http://www.cia.gov/cia/publications/factbook/rankorder/2119rank.html
[18] Another analogue is Figure A2, which portrays Zipf's law for country density, so that physical land area is accounted for; the *WDI* provides country-specific land area data back through 1965. In 2000, the cross-country correlation between the natural logarithms of population and land area was .84.
[19] Both the Ukraine (the 25th largest country in 2004, with a population of almost 48 million) and Gabon (#150, 1.4 million) are marked.
[20] Migration is also probably more permanent between cities than across countries; many immigrants eventually return to their country of origin.
[21] Urbanization rates vary dramatically across countries, so that cities do not play comparable roles in different countries. In 2000, the 90% range for the urbanization rate of the biggest 50 countries was (16%,88%), while the 50% range was (36%,76%); ranges for the entire set of countries are comparable.