UNIVERSITY OF WAIKATO

Hamilton
New Zealand

A Century of the Evolution of the Urban System in Brazil

Valente J. Matlaba, Mark J. Holmes
Philip McCann and Jacques Poot

Department of Economics
Working Paper in Economics 12/11

June 2011

Corresponding Author

Valente Matlaba
Economics Department
University of Waikato
Private Bag 3105
Hamilton, New Zealand, 3240
Email: valente.matlaba@yahoo.com.br

Mark Holmes and Jacques Poot
University of Waikato
Hamilton, New Zealand, 3240
Email: holmesmj@waikato.ac.nz
E-mail: jpoot@waikato.ac.nz

Philip McCann
Faculty of Spatial Sciences
University of Groningen
The Netherlands
E-mail: p.mccann@rug.nl

Abstract

In this paper, we study the hitherto unexplored evolution of the size distribution of 185 urban areas in Brazil between 1907 and 2008. We find that the power law parameter of the size distribution of the 100 largest urban areas increases from 0.63 in 1907 to 0.89 in 2008, which confirms an agglomeration process in which the size distribution has become more unequal. A panel fixed effects model pooling the same range of urban size distributions provides a power law parameter equal to 0.53, smaller than those from cross-sectional estimation. Clearly, Zipf's Law is rejected. The lognormal distribution fits the city size distribution quite well until the 1940s, but since then applies to small and medium size cities only. These results are consistent with our understanding of historical-political and socio-economic processes that have shaped the development of Brazilian cities.

Key words

Zipf's Law
Gibrat's Law
lognormal distribution
city size
population growth
Brazil

JEL Cssification

J11, N96, O18, R11, R12

Acknowledgements

A previous version of this paper was presented at the 57th Annual North American Meetings of the Regional Science Association International at the Grand Hyatt Hotel in Denver, November 10-13, 2010. We thank the discussant of this paper, Tatiane Almeida de Menezes, and other session participants for their comments and suggestions.

1. Introduction

Urbanization has been in recent years a key area of debate among economists (Brakman, Garretsen, van Marrewijk and van den Berg, 1999; Black and Henderson, 1999, 2003; Duranton, 2007). The urban area plays an important role in the regional economy as the spatial unit where most economic activities occur. A standard method to test whether the distribution of cities is consistent with various theories of urbanization is to check if the power law holds (Brakman, Garretsen and van Marrewijk, 2009). The power law (or Pareto distribution) holds when there is a negative loglinear relationship between the size and rank of cities, at least above a certain city size threshold.[1] This law, and specifically the slope of the loglinear relationship, is an important tool for understanding urban growth. If urban growth is a stochastic process in which every city shares the same expected growth rate and the variance of the growth rate is also the same for each city (referred to as Gibrat's Law in the literature), the distribution of city sizes is lognormal. The upper tail of the lognormal distribution closely resembles a Pareto distribution. Sometimes this Pareto distribution has a unitary slope which implies that the product of rank and size among the larger cities is constant, which is referred to as Zipf's Law (Gabaix, 1999*a*, *b*; Eeckhout, 2004, 2009).

Previous studies of the power law in various countries have three limitations. Firstly, these studies often just use cross-sectional data on cities (Rosen and Resnick, 1980; Soo, 2005; Giesen, Zimmerman and Suedekum, 2009). Other works use panel data on cities but only for short continuous time periods (Xu and Zhu, 2009; Song and Zhang, 2002).[2] Those that exploit longer time series tend to use only one observation per decade (Parr, 1985; Delgado and Godinho, 2006; Overman and Ioannides, 2001; Garmestani, Allen, Gallagher and Mittelstaedt, 2007; Moura and Ribeiro, 2006). Secondly, although the urban area (or metropolitan area) is the most appropriate geographical unit of analysis on the grounds that agglomeration externalities are better captured in this spatial unit (McCann, 2001; Brakman, Garretsen and van Marrewijk, 2009),[3] many previous studies use data on the smaller, administratively defined, cities. This biases the power law parameter downward. Thirdly, even studies that employ urban agglomeration data can be deficient if they use only cross-

---

[1] Pareto pioneered the power law in his *Cours d'Economie Politique* (1896, 1897) (The New School, 2010). However, this law was first applied in economics to the distribution of income (Simon, 1955) rather than the distribution of cities.

[2] The only exceptions are Bosker (2008, chapter 5), who analyses the distribution of the 62 largest west-German cities from 1925 to 1999; and Giesen and Suedekum (2009) who test for Gibrat's and Zipf's laws for the 71 largest west-German cities at the national level and the mostly 20 largest cities at the state level from 1975 to 1997. Both studies employ annual city data.

[3] Rosen and Resnick (1980, p. 170) note that, 'For size distribution studies, the entire metropolitan area is the most desirable choice for an urban unit as it represents an integrated economic unit. Since many workers and consumers in a city often reside in the surrounding suburbs, it seems reasonable to include these areas in the definition of the city'. Soo (2005, p. 242) adds: 'Data for agglomerations might more closely approximate a functional definition, as they typically include surrounding suburbs where the workers of a city reside'.

sectional data from which it is impossible to test whether agglomeration takes place which would be reflected in an increase in the power law parameter over time (Rosen and Resnick, 1980; Gabaix and Ioannides, 2003; Pumain and Moriconi-Ebrard, 1997; Brakman, Garretsen and van Marrewijk, 2009).

This paper overcomes these three limitations by using data on urban agglomerations at frequent intervals over a long time span. We study the size distribution of 185 urban areas in Brazil observed annually from 1907 to 2008. While there are other power law estimates for Brazil (for example, Brakman, Garretsen and van Marrewijk 2009 suggest an estimate of 0.7815 based on 193 cities), this is the first application with a long and continuous time series of urban area populations. This study aims to test whether the three aforementioned laws hold concerning the size distribution of urban areas in Brazil: namely the power law among the large cities, Zipf's law as a special case of the power law, and Gibrat's law with respect to the entire distribution.

The dataset used is unique as its construction is based on a wide range of geographical and historical information on urban activity rather than on administrative definitions of cities. The power law suggests that there is a concentration of economic activity in large agglomerations. However, such agglomerations are usually a combination of a core city together with surrounding smaller cities or towns. Our spatial unit of measurement is therefore the urban area that is consistent with urban economic theory: a single or multiple core metropolitan area that has its boundary defined by a transition from predominantly urban to predominantly rural activity. Using this definition of urban areas we confirm the power law for the 100 largest urban areas of Brazil. We also confirm the lognormal distribution for all urban areas, which is consistent with Gibrat's law. We reject Zipf's law, but find support for the increasing economic importance of urban agglomeration in the process of economic development in Brazil. We find that the power law parameter for the size distribution of increases from 0.63 in 1907 to 0.89 in 2008. A panel fixed effects model pooling the same range of urban size distributions provides a power law parameter equal to 0.53, smaller than those from cross-sectional estimation.

The paper is organized as follows. The next section discusses the background literature on the power law. Section 3 describes the data used and their sources. Section 4 briefly outlines the characteristics of the recent structural transformation in Brazil to provide the context. Section 5 discusses the empirical results. Lastly, section 6 provides concluding remarks.

2. Literature

The power law of the distribution of cities sizes is a property that applies to many distributions with fat tails. The income distribution is another socioeconomic example of a fat (right) tail distribution and it was in fact this distribution that the power law was first applied to by Pareto at the end of the 19th century. The New School (2010, paragraph 10) states:

> [Pareto] argued that in all countries and times, the distribution of income and wealth follows a regular logarithmic pattern that can be captured by the formula: Log N = log A + m log x where N is the number of income earners who receive incomes higher than x, and A and m are constants.

This law was subsequently applied to the distribution of German cities as early as 1913 by Auerbach (see Bosker, 2008; Anderson and Ge, 2005; Córdoba, 2008; Soo, 2005, 2007; Gabaix, 1999*a*, 1999*b*; Gabaix and Ioannides, 2003; Ioannides and Overman, 2003; Overman and Ioannides, 2001). Auerbach denoted the variables of the power law equation as follows: N is population size of the city with rank x, with the largest city ranked 1, the second largest city ranked 2, and so on; A and m are parameters: the former is the intercept that equals the expected value of the logarithm of the largest city and the latter is the slope which equals the power law parameter. These two parameters are usually estimated by OLS (the alternative is the Hill estimator, see Hill 1975).[4] The power law parameter m is a negative number of which the absolute value is known as α (or q) in the city size distribution literature. Zipf (1949) emphasized the special case in which $\alpha = 1$; consequently, this particular case is known as Zipf's Law (or the rank-size rule).

The estimate of α indicates the degree of city size distribution skewness. If Zipf's law does not hold there are two possibilities: i) if $\alpha > 1$, the city size distribution is more uneven and the biggest city is larger than Zipf's law predicts; ii) if $\alpha < 1$, the city size distribution is more even and the biggest city is smaller than Zipf's law expects.[5]

Of particular interest for the study of development of the urban system in a country is the change in the power law parameters over time. The change in the intercept shows the expected growth in the largest city. The change in the slope parameter suggests whether the distribution of city sizes is becoming more uneven or even. When agglomeration is becoming more important, the slope parameter increases over time and this is in fact what is found for the US (Black and Henderson, 2003). As noted in the introduction, the power law for larger cities is consistent with Gibrat's Law describing the process of urban growth. Gibrat's Law assumes independence between city growth rate and city size. When this law holds, the rank size rule is stable over time. In other words, the ratio of the largest city size to each of the other city sizes does not change over time. This urban system stability has economic implications for the distribution of employment, market areas, city innovation potential as a result of the volume of research in that city, variety of goods and services in the city, housing markets, etc. The dependence of economic aggregates of the region or country on the urban system is exactly the spirit of Christaller's and Lösch's urban hierarchy theories that connect

[4] In contrast to the original work by Pareto and by Zipf (1949), some studies put city rank on the left side of the equation and city size on the right. See Eeckhout (2004, 2009) and Bosker (2008) and the references therein.

[5] See Brakman, Garretsen and van Marrewijk (2009), chapter 7.

the complexity of economy of the urban area to the area size (Krugman, 1996; McCann, 2001; Mori, Nishikimi and Smith, 2005, 2008; Duranton, 2002, 2007).

City size distribution studies differ in sample size, degree of development of the studied country and in either rejecting or confirming Zipf's law. The literature has taken three approaches. The first approach uses cross-section data on cities to test Zipf's law and finds a power parameter either greater than 1 (Rosen and Resnick, 1980; Soo, 2005) or less than 1 (Garmestani, Allen and Gallagher, 2008; Eeckhout, 2004). The second approach makes use of a range of urban (or metropolitan) area cross-sections to comparatively reject Zipf's law (Brakman, Garretsen and van Marrewijk, 2009; Rosen and Resnick, 1980) or confirm Zipf's law (Gabaix, 1999*a*, 1999*b*; Gabaix and Ioannides, 2003; Ioannides and Overman, 2003). The third approach tests for both Zipf's law and Gibrat's law simultaneously. This approach employs panels of cities or urban (metropolitan) areas. Among studies using this approach, most reject both laws (Song and Zhang, 2002; Moura and Ribeiro, 2006; Delgado and Godinho, 2006; Pumain and Moriconi-Ebrard, 1997; Soo, 2007; Xu and Zhu, 2009; Black and Henderson, 2003; Bosker, 2008), but there is also some support (Giesen and Suedekum, 2009).

3. Data

This study uses two official data sources from Brazil: IBGE - Brazilian Institute of Geography and Statistics and IPEA - Institute of Applied Economic Research. The websites of these two institutes are sources of rich socio-economic data at national and regional levels. The administrative geographical unit we use is the municipality. We obtained municipality population data from all censuses from which municipality population data are available at this level (see Table 1); that is, covering the period from 1907 to 2008. Our sample of 185 urban areas has been built up from 1,409 municipalities in Brazil.

Table 1: Original Municipality Population Data

Year	Source	Source Obtained by
1907 to 1912	IBGE	Estimate
1920	IPEA	Census
1936 and 1937	IBGE	Estimate
1939	IBGE	Estimate
1940	IPEA	Census
1950	IPEA	Census
1960	IPEA	Census
1970	IPEA	Census
1980	IPEA	Census
1985	IBGE	Estimate
1991	IPEA	Census
1996	IPEA	Estimate
1999 to 2008	IBGE	Estimate*

*2000 is from Census.

We constructed urban areas using four steps. First, we summed the population of contiguous municipalities in 2008. We checked contiguity by means of 2009 IBGE Brazilian States Political maps.[6] We complemented these maps with Google maps.[7]

The definition of urban areas we used implied that some crossed State boundaries.[8] Therefore in some cases, an urban area is a collection of contiguous municipalities that belong to neighbouring states. The reason is essentially historical. Information of municipality history and of splits and merges of municipalities over time comes from the IBGE population data files themselves.[9] We also consulted other sources (Tenenbaum, 1996; Fausto, 1999) regarding the history of regions and settlements in Brazil

Second, we applied our urban area definition for 2008 back to 1907. We observed urban area growth both in terms of an increase in population of the municipalities and birth of new contiguous municipalities. Third, we applied smoothing to these population data under the assumption that some observed changes are inconsistent with the underlying demographic processes. This smoothing took account of neighbouring municipalities as well as temporal changes.

Fourth, we estimated urban area populations for years without data from the official sources by interpolation.[10] Comparison between the calculated and original data for years in which both types of data are available (see Table 1) suggest that our smoothing and interpolation does not distort the analysis: the correlation between the original and the adjusted data is around 0.98.

In a discussion about sample quality for the power law test, Resende (2004, p. 1547) notes the importance of using heterogeneous samples of cities. Due to data limitations our sample does not include all urban areas in Brazil, unlike Eeckhout's (2004) USA study. Yet, our sample is heterogeneous in that it involves urban areas of all sizes, in contrast with other studies that only use the largest Metropolitan Areas (such as: Gabaix and Ibragimov, 2006; Black and Henderson, 1999). This urban area size heterogeneity is achieved even though we

---

6 http://www.ibge.gov.br/

7 http://maps.google.co.nz/

8 In fact, to avoid compromising the originality of our tests for the power, Zipf and Gibrat laws for Brazil, the areas definitions created by bureaucrats and politicians have not been used. For a discussion of the importance of using functional rather than administrative urban areas, see Holmes and Lee (2010).

9 We also used www.citybrazil.com.br. However, the material on this website is essentially based on IBGE information.

10 The population of Brasilia, the capital of Brazil that was started to be built in 1956, was extrapolated back from 1960.

dropped some urban areas which appeared to have strongly oscillating populations over time. The latter may be considered outliers.

The data on urbanization in Brazil have caveats that originated in 1938 during the Getúlio Vargas Presidency in which the government elevated all municipalities to city status despite the economic structure of some municipalities not fulfilling the requirements of an urban economy. As a consequence, this overestimated urbanization in Brazil (Veiga, 2003). After an analysis of law amendments that established new municipalities in Brazil, Resende (2004, p. 1544) points to 'non-rigorous criteria for the creation of municipalities (...)'.

This urban population data problem can in recent times be solved by the use of satellite data on urban activity from EMBRAPA,[11] but these data refer only to the cross-section of Brazilian urban areas that corresponds to the 2000 population census. Although some authors advocate their use for the analysis of city size distributions (Kinoshita et al., 2008), these satellite mapping data also have limitations and are subject to criticism (Doll and Muller, 2000).

4. Structural Transformation in Brazil

4.1. Brazil's Recent Economic History

Since the arrival of the Portuguese in April 1500 and subsequent colonialization, Brazil has undergone many phases of strong social, political, economic and cultural changes. This sub-section briefly describes the main events that influenced the city size distribution from 1907 to 2008. In doing so, we define and describe 6 periods. The first period is 1907-1930, referred to as 'Development of the Republic' (Lobo, 1996, p. 426). This period is characterized by labour immigration that was needed to facilitate growth of manufacturing. Although manufacturing grew as a result of Foreign Direct Investment and exports, the economy was essentially dependent on exports of coffee. The fall in coffee prices during the 1929 depression reduced state revenue necessary for import of machinery which the industrialization policy depended on.

The second period is 1930-1945 (The Vargas Era). This period is characterized by: i) national integration policies, combination of authoritarian, totalitarian and fascist elements and the beginning of the imports substitution process (Lobo, 1996, p. 428); ii) the increase in internal migration (Fausto, 1999, p. 234); and iii) the immigration restriction policy which reduces population growth in the 1930s (Bethell, 2008; Lobo, 1996; Silva, 2008).

The third period is 1945-1964 (Democracy or 'Developmental State'). This period is marked by: i) the Kubitschek government (1956-1961) that adopted an economic policy

[11] EMBRAPA stands for Empresa Brasileira de Pesquisa Agropecuária (Brazilian Enterprise of Farming Research). These satellite data are available on http://www.urbanizacao.cnpm.embrapa.br/conteudo/base.html

inspired by Rostow's theory of take-off.[12] It concentrated investment in certain areas (Minas Gerais, São Paulo and Rio de Janeiro) where the preconditions for self-sustained growth existed (Lobo, 1996, p. 428); ii) the investment for construction of Brasilia city, inaugurated in 1960, and another migration wave from the Northeast to São Paulo (Lobo, 1996, p. 429); iii) incentives for national manufacturing intensified the imports substitution process (Fausto, 1999; Abreu, 2008).

The fourth period is 1964-1984 (Dictatorship or 'Authoritarian State'). The main characteristics are (Lobo, 1996; Fausto, 1999; Abreu, 2008*a*): i) the combination of economic stagnation and inflation ('stagflation'); ii) the annual average real income growth is 11.2% over the 'economic miracle' (1969-1974); iii) income concentration; iv) little political rights and freedom; strong regulation of the economy and creation of public institutions (1967-1974); v) oil shocks (1974-1980) causing macroeconomic instability; vi) redistribution of product that harmed the Northeast and benefited the West-Centre, North and South regions; vii) Protectionism, contractionist policies, and falling output (1981-1983).

The fifth period is the short period 1985-1989 (Democratic Transition). This period is characterised by hyperinflation and stagnation. Lastly, the sixth and most recent period is 1989-2008 (Trade liberalization and the return to Democracy (Lobo, 1996)). The main events are (Lobo, 1996; Abreu, 2008*b*; Abreu and Werneck, 2008): i) the structural reforms under Collor de Mello (1990-1992) and Itamar Franco (1992-1994) Presidencies; ii) the policies that aimed to balance inflation and unemployment were more successful since mid-1994; iii) however, as Abreu and Werneck (2008, p. 432) point out, '(…) between 1994 and 2004 per capita GDP (gross domestic product) increased [at] an average of only 0.9 percent per annum'; and (iv) despite trade liberalization, the Brazilian economy remains relatively closed over this period. Prideaux (2009, p. 16) notes that 'Brazil's imports and exports taken together were equivalent to 22% of its GDP in 2007, compared with 23% for America'.

Finally, we note that the Brazilian economy has been marked by strong State intervention throughout the 1907 to 2008 period. Politics played an active role that shaped the socio-economic structure and the city size distribution.

4.2. Urbanization in Brazil

Figure 1 shows that the population of Brazil grew from 20.3 in 1907 to 191.9 million inhabitants in 2008, which implies an annual average growth rate of 2.2%. Table 2 presents the evolution of the urban population in our sample of urban areas. The urban population defined by our sample increased from 53% of the total population in 1907 to 70% in 2008. For comparison, the urban population share was estimated by the UN Secretariat to have been

[12] This theory argues that development has mainly two stages: at the first stage the government should focus on developing regions that have the 'preconditions of self-sustained growth' (Lobo, 1996, p. 428) in order to 'take-off' the development; then, at the second stage, that development is expanded to the less developed regions. The problem with this 'selective support' of regions dependent on their development stage is that it creates regional inequality from the outset.

36.2% in 1950, increasing to 86.5% in 2008. The smallest urban area in 1907 was Goianésia do Pará with a population of 200. It remained the smallest urban area until 1945, after which Caracaraí took that place. The latter's population was 18,789 in 2008. The largest city in 1907 was Rio de Janeiro, with a population of just over 1 million. Its population increased to 4.8 million in 1960. From 1961, Sao Paulo became the largest city, with a population of close to 20 million in 2008. The average urban area population increased from 58,401 in 1907 to 730,383 in 2008.

Figure 1: Population of Brazil, 1907 to 2008

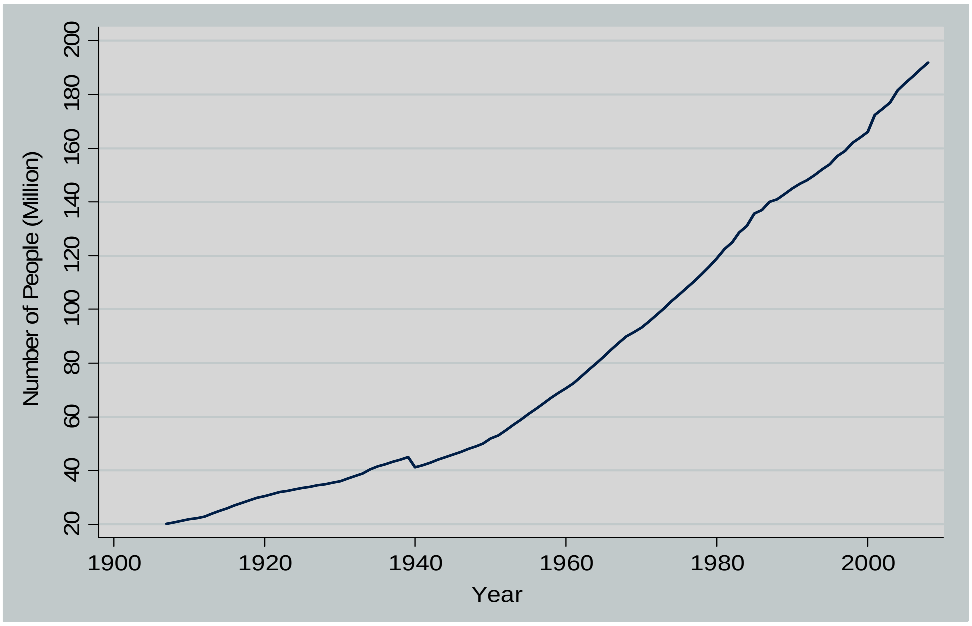


Table 2: National Population, Urban Population, and Urban Areas Sample (N=185)

Year	Total Population (1)	Total Sample Urban Population (2)	Total Sample as a Percentage of Total Population (3) = [(2)/(1)]*100	Minimum Urban Area Size	Maximum Urban Area Size	Average Urban Area Size	Percentage of Urban Population*
1907	20,253,609	10,804,332	53.35	200	1,039,082	58,401	
1910	21,819,738	11,670,719	53.49	300	1,103,057	63,084	
1920	30,559,034	14,675,734	48.02	600	1,378,865	79,328	
1930	36,000,000	18,098,944	50.27	787	1,814,562	97,832	
1940	41,169,321	20,431,303	49.63	1,200	2,203,345	110,439	
1950	51,941,078	26,507,511	51.03	869	3,137,977	143,283	36.2
1960	70,624,622	37,592,468	53.23	3,321	4,811,937	203,202	44.9
1970	93,134,846	52,516,454	56.39	4,421	8,063,414	283,872	55.8
1980	119,011,052	73,585,193	61.83	6,000	12,465,119	397,757	67.4
1990	145,000,000	93,571,199	64.53	8,577	14,800,000	505,790	74.8
2000	166,112,518	112,609,413	67.79	10,457	17,296,131	608,699	81.2
2008	191,943,158	135,120,951	70.40	18,789	19,859,740	730,383	86.5**

Notes: For the years without data on the sources, total population was assumed based on the smoothness of the population curve. The minimum area is Goianésia do Pará from 1907 to 1945 and Caracaraí from 1946 to 2008. The maximum area is Rio de Janeiro from 1907 to 1960 and São Paulo from 1961 to 2008. ** It refers to 2010.
*Source: The United Nations Secretariat, World Urbanization Prospects: The 2007 Revision, http://esa.un.org/unup.

Figures 2 and 3 visually display the evolution of the urban system in Brazil between 1907 and 2008. Clearly, the average population of urban areas 12 fold over the century. However, population growth has not been at a steady rate over the century. Over time, population growth has changed in an 'M'-shape pattern. Population growth first peaked in the 1910s, then dropped down to a low in the 1930s. After that, growth increased again until the 1950s, followed by a drop and subsequent stabilization of the growth rate by the 1990s. The second part of this 'M' pattern is consistent with the law of diminishing returns of land use. In other words, the increase of the urban area population is limited by the contiguous land area.

Figure 2: The Urban Population of Brazil, 1907

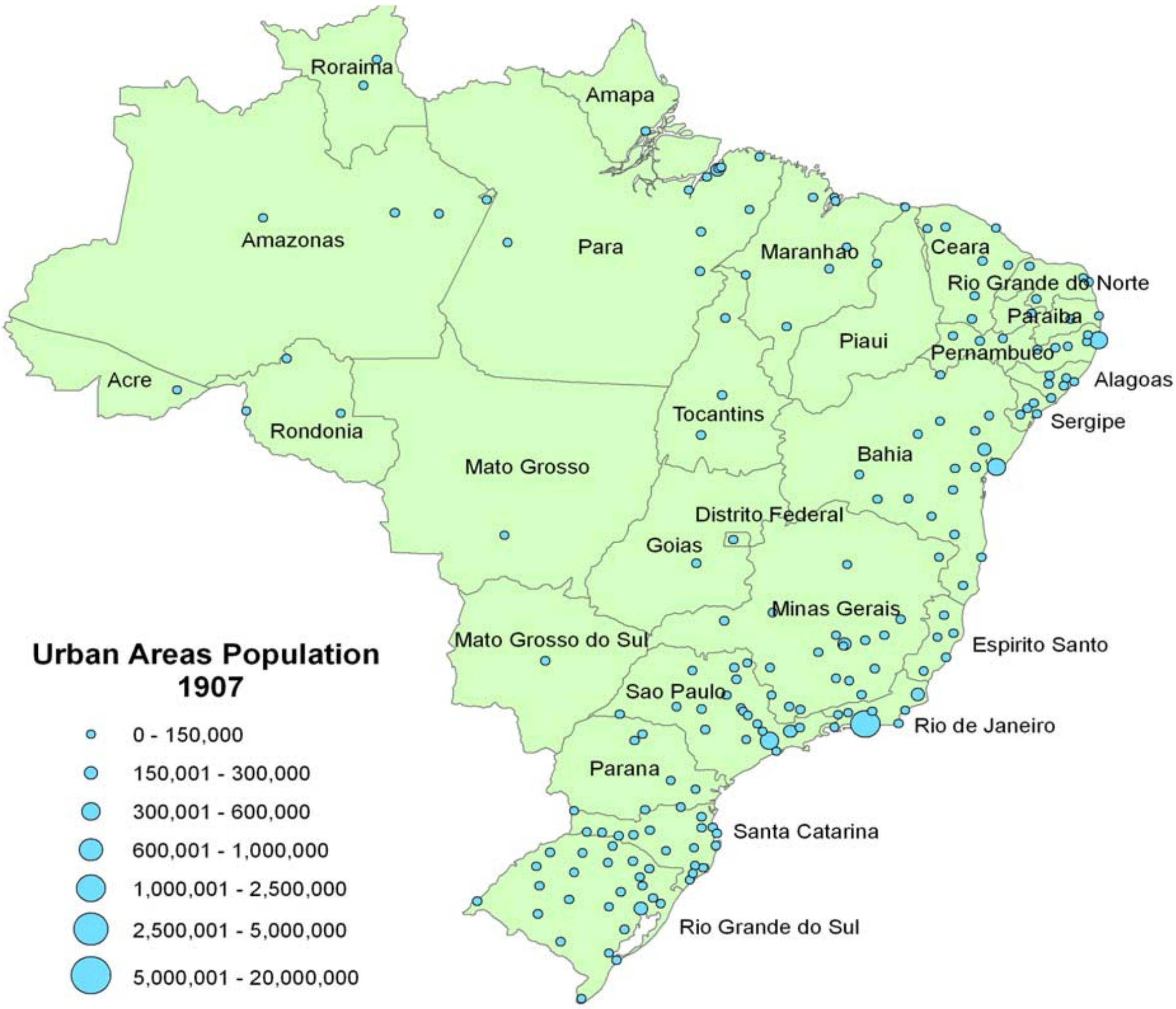
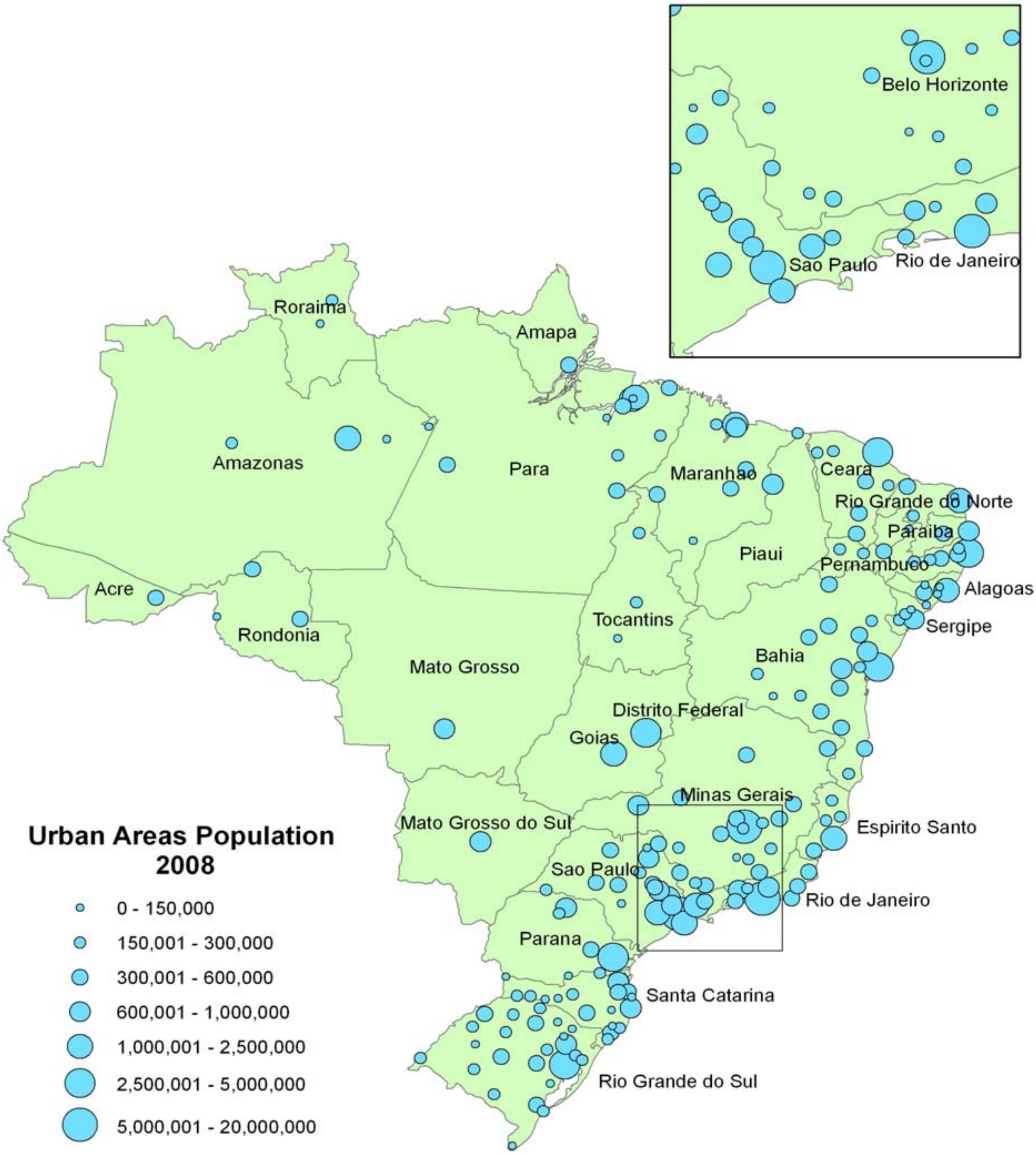


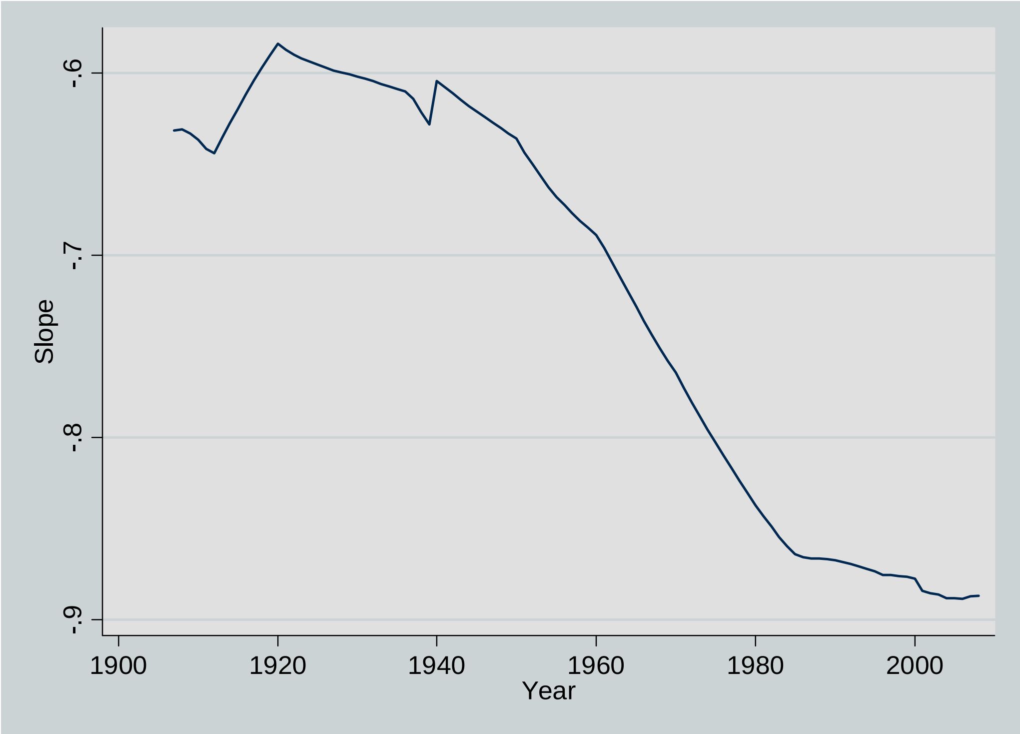
Figure 3: The Urban Population of Brazil, 2008



5. Empirical Results

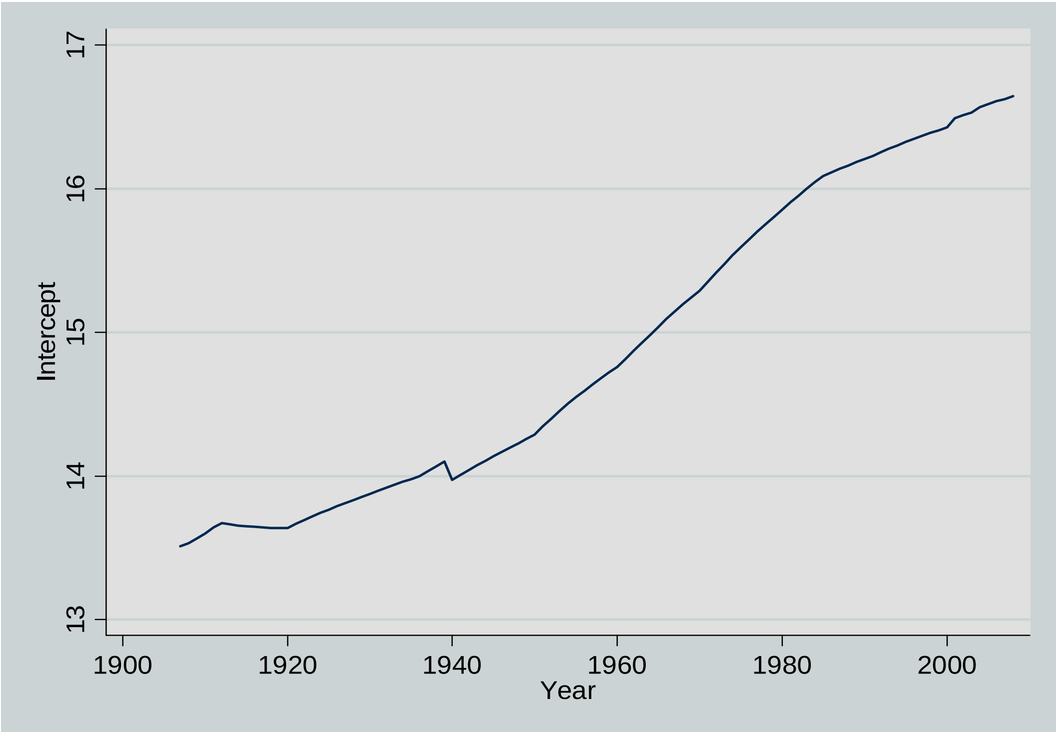
We first present OLS estimates of the power law parameters for the 100 largest urban areas. Figure 4 shows that the slope coefficient of the loglinear equation decreases from -0.63 in 1907 to -0.89 in 2008. Looking at the whole period, the figure clearly shows Gibrat's Law of proportional city growth does not hold, because the slope is not constant. Moreover, Zipf's Law does not hold either because the power law parameter is less than one. Except for 1939-1940, the city size distribution has become increasingly uneven.

Figure 4: The Slope of the Power Law Equation: Brazil, 1907 to 2008



In other words, there is a city size divergence (or a convergence to the Zipf's Law). The intercept shows a steep growth of the largest urban area: firstly, Rio de Janeiro from 1907 to 1960 (log size increases from 13.51 to 14.76), then São Paulo from 1961 to 2008 (log size grows from 14.82 to 16.64). This is shown in Figure 5.

Figure 5: The Intercept of the Power Law Equation: Brazil, 1907 to 2008



The slope behaviour can be divided into three phases. Phase I: the slope is fairly static around -0.6 from 1907 to 1939. This period included manufacturing industry development in Brazil but the economy was essentially agricultural, producing and exporting coffee - the main source of state revenue. Imports of machines to develop manufacturing were partially limited by a fall in state revenue as a result of the Great Depression (1929-1933). Labour immigration, to be employed in manufacturing, was only at the beginning.

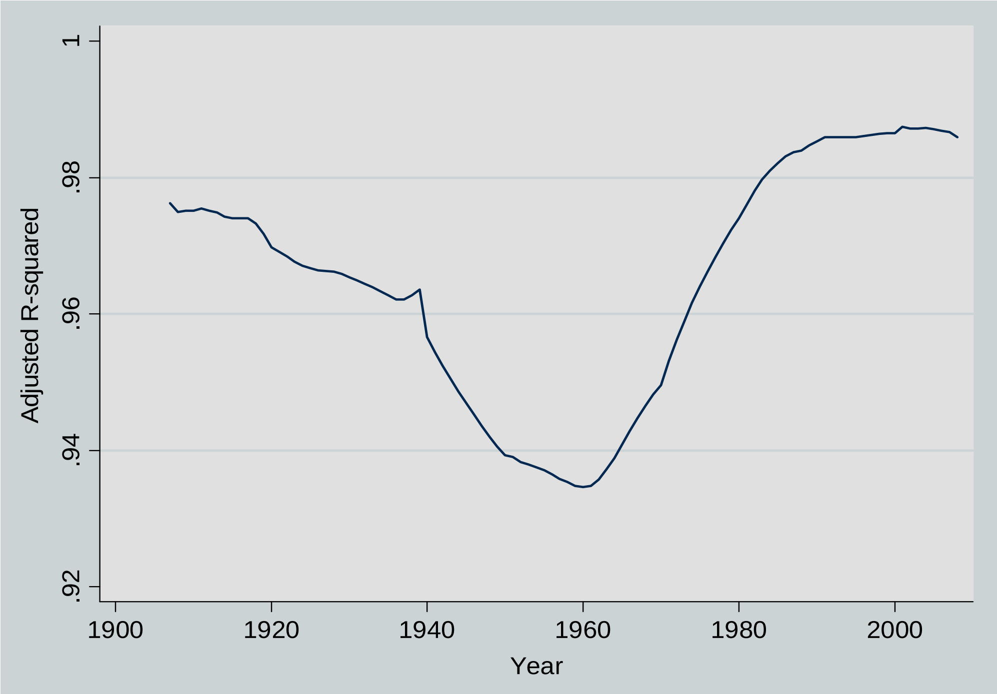
Phase II: the slope steeply falls from -0.60 to -0.86 from 1940 to 1983. In this period industrialization occurs either by import substitution or by industrialization policies that had sectoral targets. Policies favoured urban areas, immigration restriction and internal migration. Politically, this period alternates between two extremes: democracy (1945-1964) and dictatorships (1940-1945; 1964-1983). Phase III: the slope is relatively stable changing from -0.86 to -0.89 from 1984 to 2008. Trade liberalization and weak economic growth characterize this period. Politically, this period represents a return to democracy after twenty years of dictatorship.

Comparing Brazil with the USA provides an interesting contrast. The absolute value of the slope parameter based on the distribution of US Metropolitan Areas is about one for the entire twenty century. This confirms both Gibrat's Law (Krugman, 1996, p. 400) and Zipf's Law (Anderson and Ge, 2005, p. 758, footnote 1; Nitsch, 2005, p. 92; and Rossi-Hansberg and Wright, 2007, p. 598) for the US. Another interesting comparison is China. While for Brazil both laws are rejected, for China Zipf's Law is rejected before the 1979 Reforms and Gibrat's Law is accepted after the reforms (Anderson and Ge, 2005, p. 758). While simple inter-country comparisons tend to trivialise complex differences between countries, it is nonetheless clear that stages of industrialisation and development of the market economy have a major impact on the city size distributions.

The fit of the power law / Pareto distribution is very good: the Adjusted R-Squared is an 'V'-shaped curve that starts close to 0.98 in 1907 and ends close to 0.99, but with an a global minimum at 0.93 in 1960 (see Figure 6). The power law fits worst between the 1940s and the 1970s. This period coincides with Phase II of the development of the power law slope (1940-1983) during which industrialization occurred. These findings contrast with Black and Henderson (2003) using US Metropolitan Areas decadal data from 1900 to 1990. These authors found that power law fits better during the industrialization phase (e.g. R^2 is: 1900: 0.981; 1910: 0.979), and worse for the recent decades in which the US has become a services-oriented economy (e.g. R^2 is: 1980: 0.957; 1990: 0.952).[13]

[13] Industrialization of the United States economy occurred predominantly between 1880 and 1900 (http://www.britannica.com).

Figure 6: Adjusted R-Squared



As a robustness check, several cut-offs were tried with respect of defining the large urban areas. When restricting the sample to only the 10 or 40 largest urban areas, the rank-size rule holds approximately for certain sub-periods, but the slope parameter varies over time. For example, with 10 urban areas, the slope decreases from -0.96 to -1.04 from 1926 to 1939, then it increases from -1.04 to -0.95 from 1982 to 1999. Considering the top 40 urban areas, the slope decreases from -0.95 to -0.96 from 1975 to 1980, then it increases from -0.96 to -0.95 from 1981 to 1989. In both cases, outside of these time intervals the rank-size rule does not hold.

Some studies (Song and Zhang, 2002; Black and Henderson, 2003; Soo, 2007; Xu and Zhu, 2009) find that the slope increases with the movement to the upper tail of the distribution. This is also largely the case with the Brazilian data. Considering the top 10, 20, 40, and 100 urban areas and ignoring the sub-period 1907-1914 in which the absolute value of the slope of the top 100 urban areas is greater than that of the top 40, there is a monotonic increase in the slope with the movement to the upper tail from 1914 to 1953. The slope increases with the movement to the upper tail for the top 20, 40, and 100 urban areas' cut-offs for 94 (out of 102) years of our series.

To fit the lognormal distribution, we use parameters based on the sample mean and standard deviation. The sample mean is displayed in Figure 7, which shows that the natural logarithm of population of the average urban area increased from 10.33 in 1907 to 12.78 in 2008. On the other hand, the standard deviation (see Figure 8) falls from 1.31 in 1907 to 0.98 in 2008. This decline in the standard deviation of the entire urban size distribution is in contrast with the divergence among the large urban areas that was reflected in the increase in the absolute value of the power law parameter. Consequently, it is not surprising that the fit of the lognormal curve to the entire city size distribution is poor, particularly after the 1940s.

Figure 7: Lognormal Distribution Mean: Brazil, 1907 to 2008

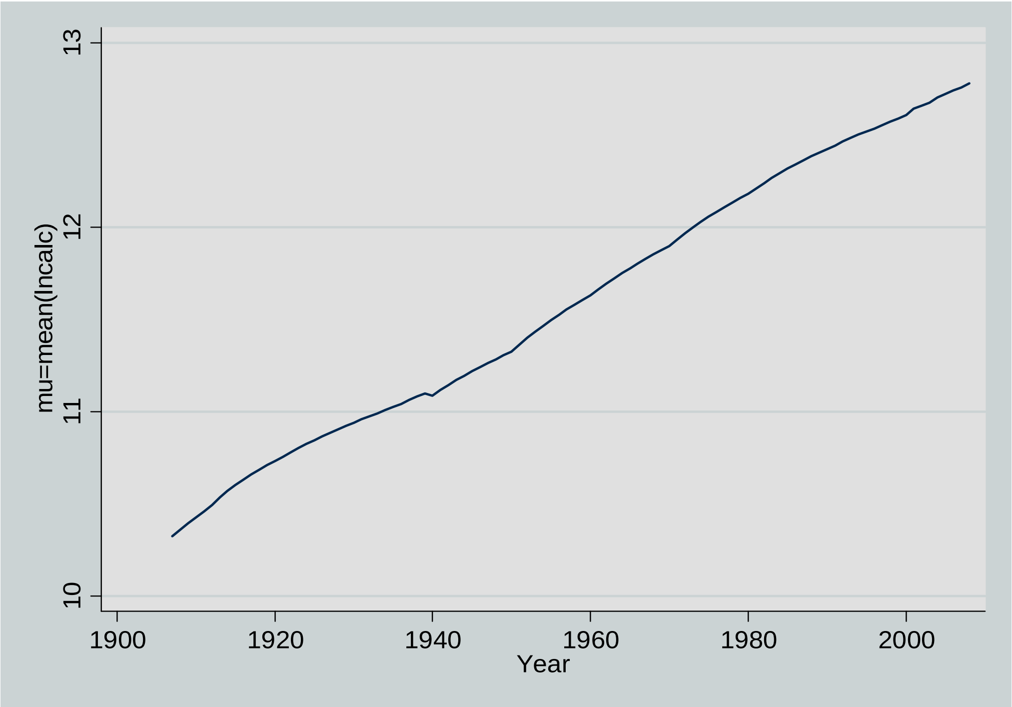
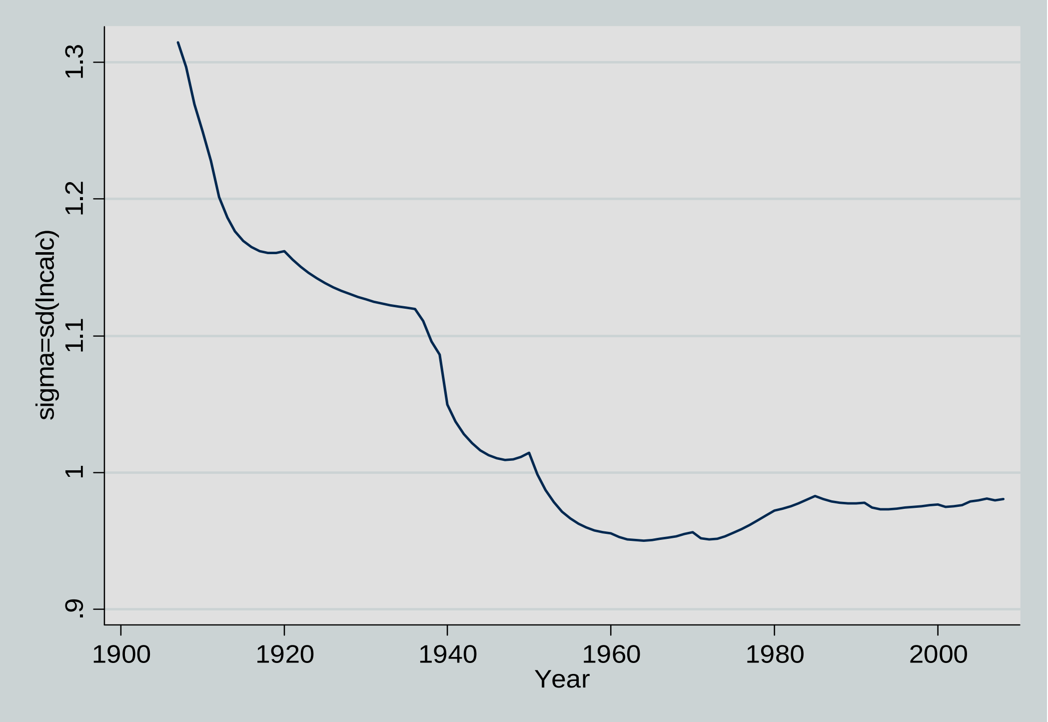


Figure 8: Lognormal Distribution Standard Deviation: Brazil, 1907 to 2008



Additionally, we estimated power law parameters for the size distribution of the 100 largest urban areas using the panel methodology of a panel model with fixed effects. The results are reported in Table 3. The estimated intercept (in logarithm) is 14.21 and the power parameter is -0.53. Clearly, the estimated panel model with fixed effects also rejected Zipf's Law. We performed a Hausman test which accepted the fixed effects model at the 5% significance level against the random effects model. The panel model also shows that the power law is a property of the cross-sectional distribution of city sizes. The model explains the 'between urban areas' variation much better (0.6745 versus 0.0810) than the 'within urban areas' (i.e. population growth of each urban area) variation. This also confirms again that Gibrat's Law did not apply to the Brazilian data.

Table 3: Panel Model with Annual Data

Variables	Fixed Effects Model Estimates
Constant	14.21 (217.11)
Logarithm of City Rank	-0.53 (-29.62)
Number of Observations	10,098
R-Squared	Within: .0810 Between: .6745 Overall: .4898

Notes: Dependent variable is logarithm of city population.
Values of *t* statistics are in brackets.

Our estimated α is consistent with the meta analysis of Nitsch (2005) who found that the power parameter is less than 1 if estimation is based on metropolitan areas (rather than city proper data) and uses post-1900 data. It is also consistent with Soo (2005) who found that the average α using urban agglomeration data is less than 1. Soo (2005) criticizes Rosen and Resnick's earlier (1980) study that suggested a value of 1 with urban agglomeration data.

Our absolute estimated power parameter from 1907 to 2008 is less than 1. Rosen and Resnick (1980, p.171) estimation for metro areas is either below 1 for Brazil, Italy and Mexico or above 1 for France and India. However, for the US it is 1, which confirms the rank size rule. Soo (2005, p.253) used urban agglomeration data of the 1990s for 26 countries to find a power parameter less than 1 for 22 cases, approximately 1 for 3 cases and higher than 1 for 1 country. Soo's (2007) estimation using urban areas in Malaysia varies from 1.08 in 1957 to 0.86 in 2000. Gangopadhyay and Basu (2009) used urban agglomerations data for India and China. For both countries, the parameter is very large and even close to 2. For example, for India, estimation suggests around 1.9 (between 1980 and 2000). For China, it is also around 1.8-2 for the same period. Brakman, Garretsen and van Marrewijk (2009, pp. 318-319) used urban agglomeration data to find power parameter greater than 1 for 9 out of 22 countries.

Our result indicates that the power parameter has increased in absolute value from 0.63 in 1907 to 0.89 in 2008. This result rejects Zipf's and Gibrat's laws, implying that urban areas' distribution in Brazil is more even than Zipf's law predicts.[14] As a consequence, the ratio of the largest urban area (São Paulo) to the second largest (Rio de Janeiro) is 1.23 for 2008. However, also for 2008, the ratio of the largest urban area to the third and the fourth largest is respectively, 3.89 and 5.35 rather than 3 and 4.

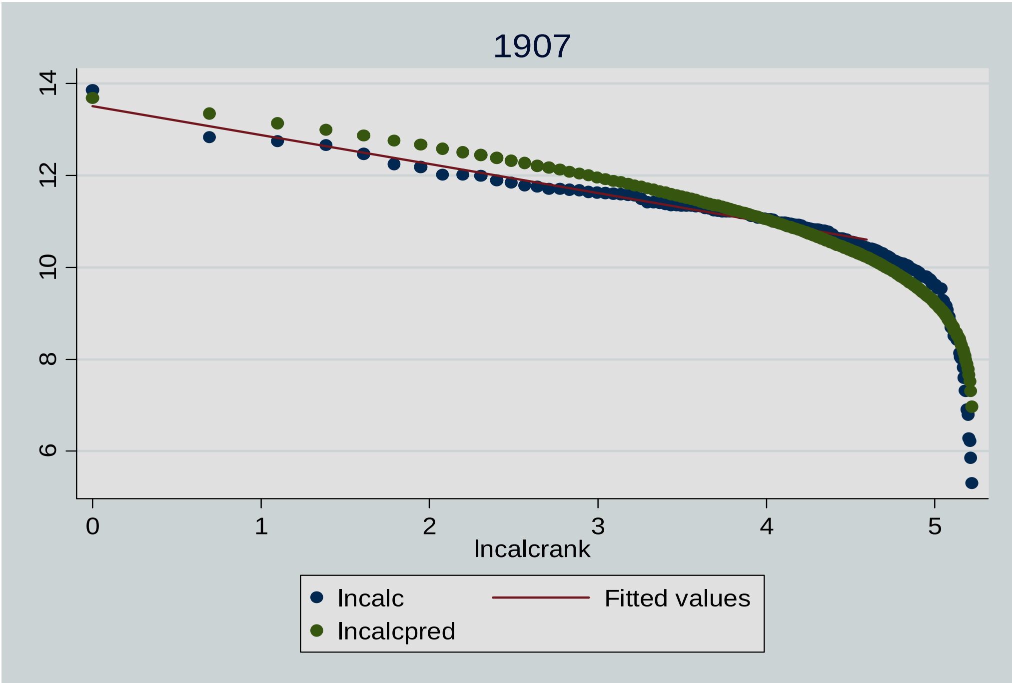
One of debates in the power law literature is whether the Pareto distribution fits the data better than the lognormal distribution. Using 2000 Census for all US cities, Eeckhout (2004, 2009) argues that both curves can fit the data equally well. On the other hand, using the same database, Levy (2009) pointed out that the lognormal curve fits better to the middle and bottom cities of the distribution, whereas the Pareto line fits the upper tail better.

The power law fits well for the 100 largest urban areas in Brazil, which is illustrated by an adjusted R^2 of at least 0.94 over the series. To test whether Eeckhout or Levy's conclusion applies to the Brazilian data, we considered cut-off of the top 100, 40, 20 and 10 urban areas. Broadly speaking, we find support for Eeckhart's claim of the equivalence of the Pareto and longnormal distributions for the larger cities for the period up to the 1950s.

For more recent years, Levy's conclusion that the lognormal of the entire distribution is inappropriate for the tail of the largest cities is also correct for Brazil. This is illustrated by comparing the fit of both distributions in 1907 and 2008, see Figures 9 and 10. Therefore, considering the entire series we conclude that both the Pareto and lognormal distributions fit well to the cities' data from 1907 to 1943 (Eeckhout, 2004, 2009; Giesen, Zimmerman and Suedekum, 2009). From 1944 to 2008 we support that the lognormal distribution fits better to the middle and bottom cities (Levy, 2009), whereas Pareto distribution describes better the very upper tail cities (Levy, 2009; Giesen, Zimmerman and Suedekum, 2009).

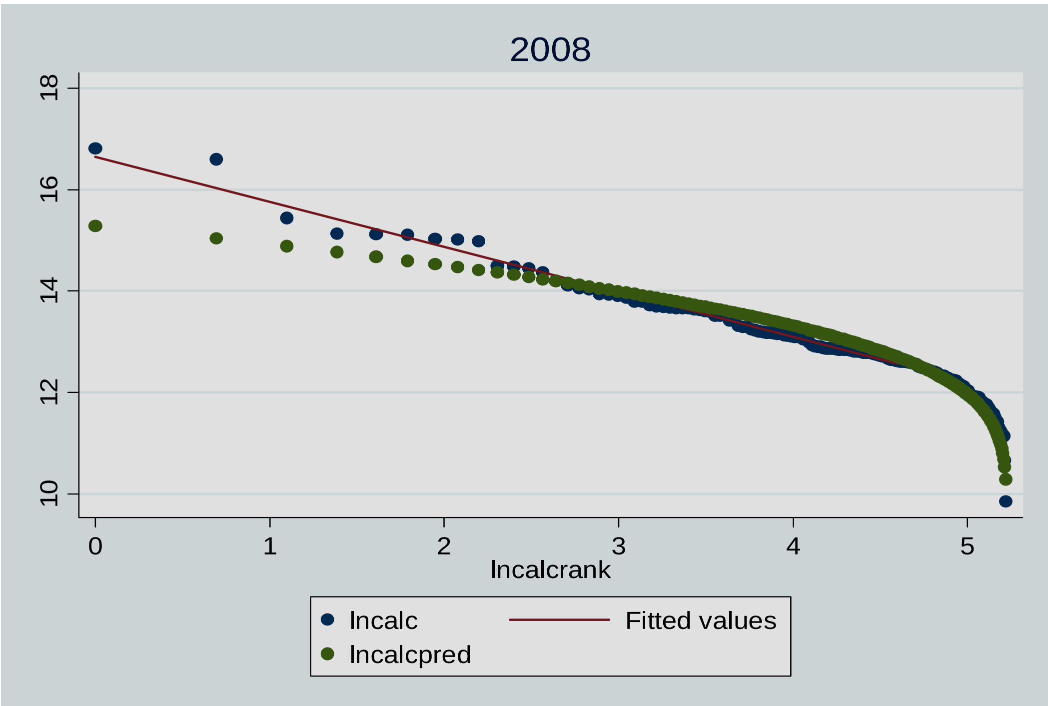
[14] Our conclusion about Gibrat's law is based on the theoretical fundamentals that relate this law to Zipf's law, rather than on regression of urban area growth rates against urban area sizes, or on the split of the urban areas into two groups – large and small – to find out if both groups have similar growth rate. However, the application of these tests would have rejected Gibrat's law given that this law holds if Zipf's law also holds (Gabaix, 1999*b*; Eeckhout, 2004).

Figure 9: Pareto and Lognormal Distributions: Brazil, 1907



Note: lncalc is the natural logarithm of population; lncalcrank is the natural logarithm of population rank; lncalcpred is the predicted logarithm of population from fitting a lognormal distribution. Fitted values refers to the power law estimated on the 100 largest cities.

Figure 10: Pareto and Lognormal Distributions: Brazil, 2008



Note: lncalc is the natural logarithm of population; lncalcrank is the natural logarithm of population rank; lncalcpred is the predicted logarithm of population from fitting a lognormal distribution. Fitted values refers to the power law estimated on the 100 largest cities.

Giesen, Zimmerman and Suedekum (2009), using 2,075 German settlements in 2006 and 25,359 USA cities in 2000, supported Eeckhout (2004) by showing that the lognormal fits these countries' city data. In our case the fit of a lognormal is not perfect because depending on the period, it either does not match some observations in the upper tail or on 'swelling' segment between the middle and the bottom of the distribution. Giesen, Zimmerman and Suedekum (2009) find that data of Germany and the USA are better described by a 'Double Pareto Lognormal' (DPLN), which is a lognormal with a Pareto fit for both the upper and the lower tails of the distribution. Since we have a sample of urban areas rather than all Brazilian urban areas in our database, we can compare these authors' findings with our study only with respect to the upper tail. Our finding differs from that of Giesen, Zimmerman and Suedekum (2009) in the sense that the Pareto fitting improvement in the upper tail is conditioned by the urban areas' cut-off and by the chosen period as shown in the comparison between our result and Levy's (2009).

6. Conclusions

This study used a unique dataset to analyze the evolution of the size distribution of urban areas in Brazil by means of a fixed sample of 185 urban areas observed annually from 1907 to 2008. Four conclusions concerning Brazil's urban system growth process can be drawn from our estimations. First, we found that the absolute value of the power law parameter (α) of the size distribution of the 100 largest urban areas increases from 0.63 to 0.89 from 1907 to 2008. Although the power law holds, Zipf's and Gibrat's laws are rejected. To verify the extent to which this result is stable irrespective of the method used, we analyzed cross-sections and time series dimensions simultaneously through employment of a panel model with fixed effects. In that case the absolute value of the power law parameter is equal to 0.53 for the 1907-2008 period, which also rejects both Zipf's and Gibrat's Laws.

Secondlywe find that α increases with the movement to the upper tail of the distribution as shown in the literature. This illustrates that the regularity that Zipf's Law states is stronger for the largest areas in Brazil. For example, the inequality among the 40 largest urban areas is higher than that observed when we consider the largest 100 urban areas, although for the former there is evidence of a decline in the absolute value of the slope parameter since the 1970s (results not shown here). Third, we find for Brazil two remarkable regularities. While the industrialization period is associated with the power parameter fall, the pre- and post-industrialization periods are related to a relatively stable parameter. The fit of the power law OLS model is worse for the industrialization phase, which is the intermediary stage of development of Brazil in which the power parameter steeply falls. However, this model performs better during the pre- and post-industrialization period of Brazil. Finally, both the Pareto and lognormal distributions describe to some extent the urban areas' size distribution during the twentieth century in Brazil. These four conclusions are consistent with theories that argue that increasing returns to scale arise as a result of agglomeration of economic activities (Brakman, Garretsen, and van Marrewijk, 2009).

This study has two caveats. First, given that a fixed sample of urban areas has been used, we were unable to evaluate the impact of birth and death of urban areas on the city size distribution. Second, we did not estimate power law parameters employing the following alternative methods: the cubic and quadratic specification equations (Rosen and Resnick, 1980; Xu and Zhu, 2009), the corrected rank-size equation proposed by Gabaix and Ibragimov (2006), the maximum-likelihood method (Kamecke, 1990; Moura and Ribeiro, 2006), the Hill estimator, or Tsallis *q*-exponential (Soo, 2007). These two caveats will be addressed in future research.

References

Abreu, M. P. (2008*a*). The Brazilian economy, 1930-1980. In L. Bethell (Ed.), *The Cambridge history of Latin America* (Vol. IX) (pp. 283-394). Cambridge, England: Cambridge University Press.

Abreu, M. P. (2008*b*). The Brazilian economy, 1980-1994. In L. Bethell (Ed.), *The Cambridge history of Latin America* (Vol. IX) (pp. 395-430). Cambridge, England: Cambridge University Press.

Abreu, M. P., and Werneck, R. L. F. (2008). The Brazilian economy, 1994-2004. In L. Bethell (Ed.), *The Cambridge history of Latin America* (Vol. IX) (pp. 431-454). Cambridge, England: Cambridge University Press.

Anderson, G., and Ge, Y. (2005). The size distribution of Chinese cities. *Regional Science and Urban Economics, 35*(6), 756-776.

Bethel, L. (2008). Politics in Brazil under Vargas, 1930-1945. In L. Bethell (Ed.), *The Cambridge history of Latin America* (Vol. IX) (pp. 3-86). Cambridge, England: Cambridge University Press.

Black, D., and Henderson, J. V. (1999). A theory of urban growth. *Journal of Political Economy, 107*(2), 252-284.

Black, D., and Henderson, J. V. (2003). Urban evolution in the US. *Journal of Economic Geography*, *3*(4), 343-372.

Bosker, M. (2008). *The empirical relevance of geographical economics*. Unpublished PhD thesis, Utrecht University, Utrecht, The Netherlands.

Brakman, S., Garretsen, H., van Marrewijk, C., and van den Berg, M. (1999). The return of Zipf: Towards a further understanding of the rank-size distribution. *Journal of Regional Science, 39*(1), 739–767.

Brakman, S., Garretsen, H., and van Marrewijk, C. (2009). *The new introduction to geographical economics.* Cambridge, England: Cambridge University Press.

Córdoba, J-C. (2008). On the distribution of city sizes. *Journal of Urban Economics, 63*(1), 177–197.

Delgado, A. P., and Godinho, I. M. (2006). *Long term evolution of the size distribution of Portuguese cities.* Working Paper #230, Faculty of Economics, University of Porto.

Doll, C., and Muller, J-P (2000). *Mapping Anthropogenic activities from Earth observation data* [PowerPoint Slides]. The Center for International Earth Science Information Network (CIESIN), Columbia University.

Duranton, G. (2002). *City size distributions as a consequence of the growth process*. Available from ideas.repec.org/p/cep/cepdps/0550.html.

Duranton, G. (2007). Urban Evolutions: The Fast, the Slow, and the Still**.** *American Economic Review*, *97*(1), 197-221.

Eeckhout, J. (2004). Gibrat's law for (all) cities. *American Economic Review*, *94*(5), 1429–1451.

Eeckhout, J. (2009). Gibrat's law for (all) cities: Reply. *American Economic Review, 99*(4), 1676–1683.

Encyclopaedia Britannica website: www.britannica.com <Accessed June, 2010>.

Fausto, B. (1999). *A concise history of Brazil.* (A. Brakel, Trans.). Cambridge, England: Cambridge University Press. (Original work published 1998).

Gabaix, X. (1999*a*). Zipf's law and the growth of cities. *American Economic Review* (Papers and Proceedings), *89*(2), 129–132.

Gabaix, X. (1999*b*). Zipf's law for cities: An explanation. *The Quarterly Journal of Economics, 114*(3), 739-767.
Gabaix, X., and Ioannides, Y. M. (2003). *The evolution of city size distributions*. Working Paper #310, Department of Economics, Tufts University.
Gabaix, X., and Ibragimov, R. (2006). *Log(Rank-1/2): A simple way to improve the OLS estimation of tail exponent*. Working Paper #2106, Harvard Institute of Economic Research.
Gangopadhyay, K., and Basu, B. (2009). City size distributions for India and China. *Physica A: Statistical Mechanics and its Applications, 388*(13), 2682-2688.
Garmestani, A. S., Allen, C. R., Gallagher, C. M., and Mittelstaedt, J. D. (2007). Departures from Gibrat's law, discontinuities and city size distributions. *Urban Studies, 44*(10), 1997-2007.
Garmestani, A. S., Allen, C. R., and Gallagher, C. M. (2008). Power laws, discontinuities and regional city size distributions. *Journal of Economic Behavior and Organization*, *68*(1), 209-216.
Giesen, K., and Suedekum, J. (2009). *Zipf's law for cities in the regions and the country*. Working Paper #3928, The Institute for the Study of Labor (IZA).
Giesen, K., Zimmerman, A., and Suedekum, J. (2009). *The size distribution across all cities – Zipf, lognormal, double Pareto, or what?* Paper presented at the North American Regional Science Council, San Francisco. Google maps website. http://maps.google.co.nz/ <Accessed through second semester 2009>.
Hill, B.M. (1975) A simple approach to inference about the tail of a distribution. *Annals of Statistics*, 3, 1163-1174.
Holmes, T. J., and Lee, S. (2010). Cities as six-by-six-mile squares: Zipf's law? In E. L. Glaeser (Ed.). *Agglomeration Economics* (pp. 105-131). Cambridge, MA: University of Chicago Press.
IBGE (2010). Brazilian Institute of Geography and Statistics. www.ibge.gov.br.
Ioannides, Y., and Overman, H. G. (2003). Zipf's law for cities: An empirical examination. *Regional Science and Urban Economics, 33*(2), 127–137.
IPEA (2010). Institute of Applied Economic Research. www.ipea.gov.br.
Kamecke, U. (1990). Testing the rank size rule hypothesis with an efficient estimator. *Journal of Urban Economic*s, *27*(2), 222-231.
Kinoshita, T., Kato, E., Iwao, K., and Yamagata, Y. (2008). Investigating the rank-size relationship of urban areas using land cover maps, *Geophysical Research Letters*, *35*, L17405.
Krugman, P. (1996). Confronting the urban mystery. *Journal of the Japanese and International Economies, 10*(4), 399–418.
Levy, M. (2009). Gibrat's law for (all) cities: Comment. *American Economic Review, 99*(4), 1672–1675.
Lobo, E. M. L. (1996). Brazil since 1889. In B. A. Tenenbaum (Ed.). *Encyclopedia of Latin American history and culture* (pp. 424-432). New York: Simon and Schuster Macmillan.
McCann, P. (2001). *Urban and regional economics*. New York: Oxford University Press.
Mori, T., Nishikimi, K., and Smith, T. E. (2005). A divergence statistic for industrial localization. *The Review of Economics and Statistics, 87*(4), 635-651.
Mori, T., Nishikimi, K., and Smith, T. E. (2008). The number-average size rule: A new empirical relationship between industrial location and city size. *Journal of Regional Science, 48*(1), 165–211.
Moura, N. J., and Ribeiro, M. B. (2006). Zipf law for Brazilian cities. *Physica A,* 367, 441-448.
The New School (2010). *Vilfredo Pareto, 1848-1923*. History of Economic Thought, The New School. http://homepage.newschool.edu/het/
Nitsch, V. (2005). Zipf zipped. *Journal of Urban Economics, 57*(1), 86–100.
Overman, H. G., and Ioannides, Y. M. (2001). Cross-sectional evolution of the US city size distribution. *Journal of Urban Economics, 49*(3)**,** 543-566.
Parr, J. B. (1985). A note on the size distribution of cities over time. *Journal of Urban Economics, 18*(2), 199-212.
Prideaux, J. (2009, November 14). Getting it together at last: A special report on business and finance in Brazil. *The Economist*, 2-16.
Pumain, D., and Moriconi-Ebrard, F. (1997). City size distributions and metropolisation. *Geojournal*, *43*(4), 307-314.

Resende, M. (2004). Gibrat's law and the growth of cities in Brazil: A panel data investigation. *Urban Studies, 41*(8), 1537-1549.
Rosen, K. T., and Resnick, M. (1980). The size distribution of cities: An examination of the Pareto law and primacy. *Journal of Urban Economics, 8*(2), 165–186.
Rossi-Hansberg E., and Wright M. (2007). Urban structure and growth. *Review of Economic Studies, 74*(2), 597-624.
Silva, N. do V. (2008). Brazilian society: Continuity and change, 1930-2000. In L. Bethell (Ed.), *The Cambridge history of Latin America* (Vol. IX) (pp. 455-544). Cambridge, England: Cambridge University Press.
Simon, H. (1955). On a class of skew distribution functions. *Biometrika, 42*(3/4), 425-440.
Song, S., and Zhang, K. H. (2002). Urbanisation and city size distribution in China. *Urban Studies, 39*(12), 2317–2327.
Soo, K. T. (2005). Zipf's Law for cities: A cross-country investigation. *Regional Science and Urban Economics*, *35*(3), 239–263.
Soo, K. T. (2007). Zipf's law and urban growth in Malaysia. *Urban Studies, 44*(1), 1–14.
Tenenbaum, B. A. (1996) (Ed.). *Encyclopedia of Latin American history and culture*. New York: Simon and Schuster Macmillan.
Veiga, J. E. (2003). Cidades imaginárias: O Brasil é menos urbano do que se calcula *[Imaginary cities: Brazil is less urban than it is calculated]* (2nd ed.). Campinas, São Paulo, Brazil: Editora Autores Associados.
www.google.co.nz/ <Accessed through second semester 2009>.
Xu, Z., and Zhu, N. (2009). City size distribution in China: Are large cities dominant? *Urban Studies*, *46*(10), 2159–2185.
Zipf, G. K. (1949). *Human behavior and the principle of least effort: An introduction to human ecology.* Cambridge, MA: Addison-Wesley.