《人文及社會科學集刊》
第二十五卷第三期（102/9）, pp. 525-550

Zipf's Law and City Size Distribution in China

Chun-hung A. Lin
Professor
Department of Industrial Economics, Tamkang University

Mei-chun Cheng
Associate Research Fellow
Taiwan Institute of Economic Research

Chih-hai Yang
Professor
Department of Economics, National Central University

ABSTRACT

China's experience with high economic growth has undergone dramatic change following the reforms initiated in 1978, which not only introduced economic incentives, but also encouraged urbanization. Due to the rapid urbanization that has taken place, it is particularly worth investigating the size distribution of Chinese cities. This study is thus an attempt to examine the size distribution of China's cities over the 1984-2008 period. The results indicate that the size distribution of Chinese cities is more equal than would be predicted by Zipf's law. We also find that the size distribution of cities is more even for larger cities than for smaller ones. According to our results, a possible explanation for the more even distribution of large cities relative to the smaller ones is related to China's government expenditures and industrial structure. Finally, we find that economic growth promotes population concentration, and that lower transportation costs will promote population concentration.

Key Words: city size distribution, Zipf's law, Pareto exponent, China

---

Received: September 13, 2012; Accepted: March 27, 2013

I. Introduction

Urbanization is a complex process by which a country's organized communities become larger, more specialized and more efficient. Urbanization may result from many factors of a demographic, economic, technological or environmental nature. In the field of economics, urbanization creates a great number of job opportunities. It encourages surplus workers in rural areas to relocate to urban areas, thus resulting in urbanization and the accompanying economic development. Urbanization brings with it several consequences—both adverse and beneficial. It strongly impacts the social population and the environment they are living in. Thus, urbanization is a highly relevant phenomenon that cannot be neglected.

Ever since the reforms that were initiated in 1978, China has experienced high economic growth which has not only introduced economic incentives and opened the economy to foreign trade and investment, but has also encouraged urbanization. This can be seen in Table 1. Whilst China's overall population has increased 1.38 times from 962.59 million in 1978 to 1,328.02 million in 2008, its urban population has increased 3.5 times, from 172.45 million in 1978, to 606.67 million in 2008. In 1978 when the reforms began, China had 193 cities with only 17.92% of the total population living in urban areas. By 2008, the number of cities had increased to 655, with 45.68% of the total population being located in urban areas. China implemented a scheme for the development of the urban areas, which was based on the guideline of "strict control of the large-sized cities, reasonable development of the medium-sized cities, and aggressive development of the small-sized cities." In the 1980s, the medium-sized cities and small-sized cities grew rapidly, and many satellite towns developed near the large-sized cities.

Rapid urbanization can also be observed by the increase in the number of cities. Chinese cities[1] are classified into five size categories according to their total population. Table 2 shows the definitions and number of cities in each category and the difference between the two periods, 1978 and 2008. As shown in Table 1, the number of urban cities in China increased from 193 in 1978 to 655 in 2008. Having experienced an increase in each category, in 1978 China had 10 super large-sized cities, 19 very large-sized cities, 35 large-sized cities, 80 medium-sized cities and 49 small-sized cities. These numbers increased to 41, 81, 118, 151 and 264 in 2008.

1 There are three different administrative levels of cities in the Chinese urban system: county-level cities, prefecture-level cities and central municipalities. Small settlements, such as towns and villages or lower administrative levels, are not treated as cities.

Table 1. Urban population in China (1,000,000 people)

Year	Total population	Urban population	Percentage urban	Number of cities
1978	962.59	172.45	17.92	193
1980	987.05	191.40	19.39	223
1984	1034.57	240.17	23.01	295
1985	1058.51	250.94	23.71	324
1986	1075.07	263.66	24.52	348
1987	1093.00	276.74	25.32	382
1988	1110.26	286.61	25.81	430
1989	1127.04	295.40	26.21	449
1990	1143.33	301.95	26.41	466
1991	1158.23	312.03	26.94	477
1992	1171.71	321.75	27.46	517
1993	1185.17	331.73	27.99	570
1994	1198.50	341.69	28.51	622
1995	1211.21	351.74	29.04	640
1996	1223.89	373.04	30.48	665
1997	1236.26	394.49	31.91	666
1998	1247.61	416.08	33.35	664
1999	1257.86	437.48	34.78	665
2000	1267.43	459.06	36.22	659
2001	1276.27	480.64	37.66	657
2002	1284.53	502.12	39.09	641
2003	1292.27	523.76	40.53	653
2004	1299.88	542.83	41.76	655
2005	1307.56	562.12	42.99	656
2006	1314.48	577.06	43.90	656
2007	1321.29	593.79	44.94	656
2008	1328.02	606.67	45.68	655

Source: *China Statistical Yearbook*, 1999 and 2009 (NBS, 1985–2009), and authors' calculations.

Note: Total population refers to the total number of people alive at a certain point of time within a given area. The annual statistics on total population is taken at midnight, the 3lst of December.

Table 2. Number of cities for different sizes in 1978 and 2008

Size	1978	2008	Difference
Total number of cities	193	655	462
Super large-sized cities (above 2 million persons)	10	41	31
Very large-sized cities (between 1-2 million persons)	19	81	62
Large-sized cities (0.5-1 million persons)	35	118	83
Medium-sized cities (0.2-0.5 million persons)	80	151	71
Small-sized cities (less than 0.2 million persons)	49	264	215

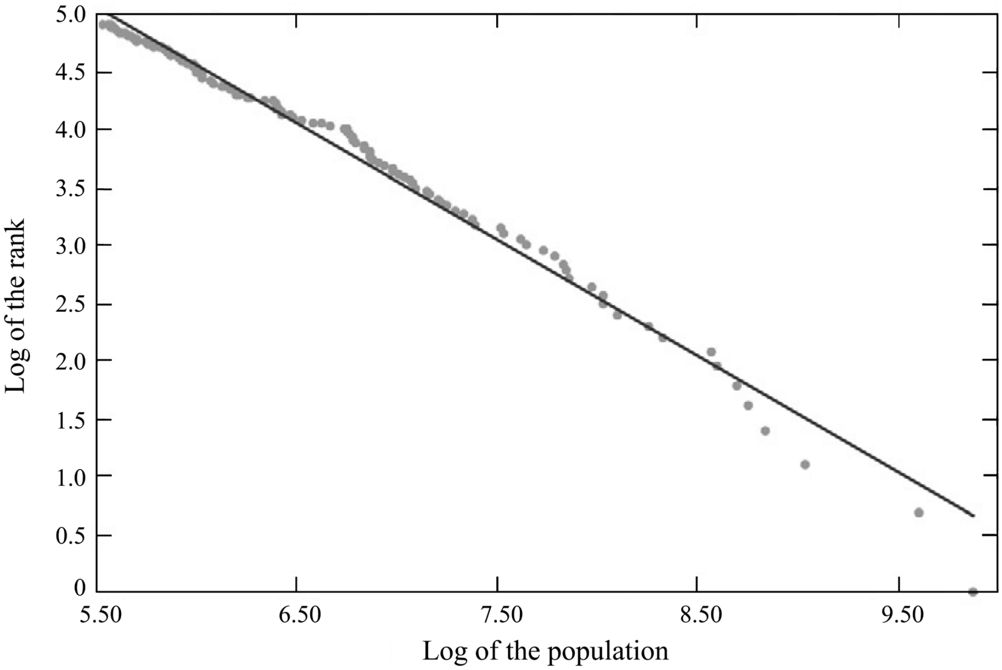
Source: *China City Statistical Yearbook* (NBS, 2009).

Tables 1 and 2 indicate that China has experienced rapid urban growth over the last 30 years.

An important consequence of urbanization is the change in the size distribution of cities. The central purpose of this study is to investigate the trend of the size distribution of cities in China and factors that may contribute to the change in this distribution. Auerbach (1913) first proposed that the size distribution of cities could be represented by a Pareto distribution. Subsequently, Singer (1936) empirically analyzed the size distribution of cities, followed by Zipf's (1949) construction of the theoretical foundation. In empirical studies, when cities are ordered by population size, regressing the logarithm of their rank on the logarithm of their size yields a slope coefficient, i.e., the Pareto exponent. The Pareto exponent is employed as a measure of the population concentration of cities of different sizes. When the Pareto exponent equals 1, we state that it satisfies Zipf's law. To get an idea of how well this works, consider the United States and how its cities are ordered in terms of the size of their respective populations. According to Gabaix (1999), we show the plot for the log of metropolitan area size against the log of rank (i.e., New York = 1, Los Angeles=2, etc.) for the 130 such areas listed in the *Statistical Abstract of the United States, 1993* in Fig. 1. The regression leads to the result that the slope of the fitted line is equal to −1.003 and the R-squared is very close to 1, with a value of 0.986.

The related empirical studies in the literature can be classified into two major categories. The first one is based on examining a single country, and the major issue is to investigate whether the size distribution of cities follows Zipf's law (for example, Krugman, 1996 and Soo, 2007). The second category discusses the Pareto exponent across countries and investigates the determinants of the Pareto exponent (for example, Rosen and Resnick, 1980; Soo, 2005; Nitsch, 2005). However, there are only a few studies that discuss the evolution of urban systems over time and

Fig. 1. Plot of 130 American metropolitan areas listed in the *Statistical Abstract of the United States, 1993* (following Gabaix, 1999)



that examine the factors which may influence the value of the Pareto exponent in a single country.

Since China experienced rapid urbanization in the year 1978, it is worth investigating the size distribution of Chinese cities and discussing how the economic and geographical factors affect the size distribution of the cities. Recently, several studies have examined the sizes of Chinese cities. For example, Song and Zhang (2002) used data for Chinese cities for only the years 1991 and 1998 to investigate the size distribution of Chinese cities and the changes that had taken place. However, there are only a few studies (for example, Anderson and Ge, 2005) that investigate the evolution of the size distribution of Chinese cities.

This paper has two purposes. First, we present the empirical results on the Pareto exponent of China from 1984 to 2008. Second, we investigate the explanations of the variation in the Pareto exponent. The remainder of the paper is organized as follows: Section 2 presents an overview of the previous literature and examines the size distribution of cities with respect to the Pareto distribution. Section 3 presents the empirical model, discusses the variables explaining the variation in the Pareto exponent, and provides information regarding the data set. Section 4 discusses the empirical results. The last section concludes.

II. Related literature

In 1913, Auerbach noted that the size distribution of cities could be represented by a Pareto distribution. A number of studies have, since the time of Auerbach, investigated whether the size distribution of cities follows a Pareto framework. Among them, the first systematic empirical analysis of the size distribution of cities was conducted by Singer (1936). He concluded that the size distribution of cities follows a Pareto distribution, and the Pareto exponent can serve as a concise measure of population concentration. Zipf (1949), in his construction of a theoretical foundation, proposed that there are two driving forces behind urban development: (1) the force of diversification that causes population dispersal; (2) the force of unification, which enables population concentration. The Pareto exponent represents the ratio of magnitude of the force of diversification divided by the magnitude of the force of unification. If the force of diversification and force of unification are virtually equal, we refer to it as Zipf's law or the rank-size rule. In other words, low values of the Pareto exponent imply that the city-size distribution has few large cities and many small cities, which are less equally sized, while high values imply a more equally sized city distribution.

Several studies have attempted to derive Zipf's law (i.e., where the Pareto exponent is equal to 1) theoretically for city-size distributions. Gabaix (1999), for instance, indicates that if cities grow randomly with the same expected growth rate and the same variance (Gibrat's law), the limit distribution will converge to Zipf's law. Giesen and Südekum (2011) attempt to analyze whether Zipf's law holds for the size distribution of German cities. They find that Gibrat's law not only holds at the national level in Germany, but also tends to hold in almost every region regardless of type. Moreover, they also find that both national and regional levels in Germany tend to follow Zipf's law. This is consistent with Gabaix's (1999) theory. However, Gan et al. (2006) find that Zipf's law is a statistical phenomenon, and therefore, it does not require an economic theory to determine city-size distributions. The Pareto exponent has traditionally been estimated using the OLS estimator. However, Gabaix and Ioannides (2004), using Monte Carlo simulations, found that this estimator was biased in small samples and proposed using Hill's estimator (Hill, 1975). Gabaix and Ibragimov (2011) showed that a simple and practical adjustment to the OLS estimator could correct the bias in small samples.

On the other hand, the related empirical studies in the literature can also be classified into two major categories. The first category was to investigate Zipf's law on a single-country basis, for example, Krugman (1996) and Soo (2007). Krugman

(1996) used 130 metropolitan areas listed in the *Statistical Abstract of the United States, 1993* and ranked them by population size before regressing the log of rank on the log of population to obtain a slope equal to −1.003 (std. dev. 0.010) and $R^2=0.986$. He repeated the analysis for populations greater than 100,000, 250,000, 500,000, and 1 million, respectively, for the three years, 1890, 1940, and 1990. The rank-size relationship is nearly linear in logarithmic form and the slope is very close to −1. Soo (2007) performed a test of Zipf's law using data for Malaysian cities from 1957, 1970, 1980, 1991 and 2000 population censuses based on two estimation methods—OLS and the Hill estimator. He found that, at the upper tail, cities were becoming more equal in size to each other; that is, the Pareto exponent is increasing over time, in contrast to the falling exponent when considering the entire distribution.

The second category discusses the Pareto exponent for cross-country data and investigates the factors that may influence the value of the Pareto exponent, that is, the factors that cause population to disperse or alternatively concentrate. As in many other cross-country studies, including Rosen and Resnick (1980), Soo (2005) and Nitsch (2005), it has been pointed out that the Pareto exponent is significantly different from 1. Parr (1985) has shown evidence from 12 countries that the values of the Pareto exponent tend to exhibit a U-shaped trend over time. The key empirical article in comparisons across countries is Rosen and Resnick (1980), whose results show that 32 out of 44 countries have Pareto exponents greater than unity and the mean Pareto exponent is 1.136. They conclude that city sizes in most countries are more evenly distributed than would be predicted by Zipf's law. Their findings are largely confirmed by Soo (2005), who uses a new dataset ranging from 1972 until 2001 for 73 countries that is combined with two estimation methods—OLS and the Hill estimator. With either estimator, Zipf's law is rejected for 53 out of 73 countries using OLS, and for 30 out of 73 countries using the Hill estimator. The mean of the Pareto exponent is approximately 1.11. As for the Hill estimator, the mean of the Pareto exponent is 1.167, which is statistically different from the mean for the OLS estimator. Nitsch (2005) applies meta-analytic procedures to summarize the empirical findings on Zipf's law for cities. The meta-analysis combines 515 estimates from 29 studies, covering a wide range of different territories and time periods. His findings are that the Pareto exponent is on average larger than 1, and close to 1.1, also implying that cities are on average more evenly distributed than suggested by Zipf's law. This finding is remarkably similar to the results from the cross-country studies by Rosen and Resnick (1980) and Soo (2005).

The Pareto exponent is employed as a measure of population concentration among cities of different sizes. If on the one hand obtaining the value for the Pareto exponent for different countries is interesting in itself, it ensures that there is great interest in investigating the factors that may influence the value of the Pareto exponent on the other. For instance, Rosen and Resnick (1980) found that the Pareto exponent is positively related to per capita GNP and total population, but negatively related to railway mileage density and land area. Subsequently, a cross-country study by Alperovich (1993) using values of the Pareto exponent from Rosen and Resnick (1980) found that it is positively related to per capita GNP, population density, and land area, and negatively related to the government share of GDP, and the share of manufacturing value added in GDP. Recently, Soo (2005) also indicated that the variation in the values of the Pareto exponent is better explained by political economy variables than by economic and geographical variables.

Recently, Song and Zhang (2002), using data for Chinese cities in 1991 and 1998, investigated the size distribution of Chinese cities and its changes. They found that Chinese cities became more evenly distributed in the 1990s. This finding suggests that the sizes of new cities made cities in the full sample appear more evenly distributed. They also proposed some economic and institutional factors that may affect the urban system and the patterns of urban growth. Differing from previous investigations which were based on fixed sample data, Peng (2010) used rolling sample regression methods in which the sample was changing with the truncated point. He found that the Pareto exponent was almost monotonically decreasing in the truncated point for the rolling rank regressions; the mean estimated coefficient was 0.84 for the full dataset, which is not so far from 1.

Urbanization is a complex process by which a country's organized communities become larger, more specialized and more efficient. Initially, the process of production activities and economic development create a tribe. Then tribes evolve into agricultural villages. After the industrial revolution, a great number of job opportunities were created in urban areas encouraging workers in rural areas to relocate to urban areas, thus resulting in urbanization and the accompanying economic development. As more and more economic activities are concentrated in cities, it will promote population concentration in urban areas, and therefore urban growth.

With regard to the city size distribution, Gabaix (1999) indicates that if cities grow randomly with the same expected growth rate and the same variance (Gibrat's law), the limit distribution will converge to Zipf's law. However, urban development has a close relationship with the size of the country, natural resources, population and economic growth. Thus, a number of studies used several economic variables to investigate what factors may influence the city-size distributions, namely what factors cause population to disperse or alternatively concentrate (Rosen and Resnick, 1980; Alperovich, 1993; Soo, 2007). They found city-size distributions could be explained by per capita GNP, population density, government share of GDP, level of industrialization, and railway mileage density. As mentioned above, Zipf's law is not only statistical but also has economic implications.

As in many other studies, this paper acknowledges the Pareto distribution as being a good approximation to the size distribution of cities in the real world. Moreover, many theoretical works have attempted to derive Zipf's law for city-size distributions. There are many empirical studies examining the size distribution of cities, of which a number of studies investigate cross-country or single country comparisons over different short periods. However, there have been very few studies on the size distribution of cities in individual countries using long-term time series data. Since China has faced rapid urbanization beginning in the 1980s, it is worth investigating the size distributions of Chinese cities and discussing how economic and geographic factors affect these distributions.

III. Model and data

The form of the size distribution of cities as first suggested by Auerbach (1913) is based on the following Pareto distribution:

$$y = Ax^{-\alpha}, \tag{1}$$

where x is the population of a city; y is the rank of the city as cities are ordered from the largest to the smallest; A is the constant term; and α is the Pareto exponent. A popular way to estimate the Pareto exponent is to run the following OLS log-log rank-size regression:

$$\log y = \log A - \alpha \log x. \tag{2}$$

According to Singer (1936) and Zipf (1949), we know that the Pareto exponent, α ($\alpha > 0$), is employed as a measure of population concentration among cities of different sizes. Larger values of the Pareto exponent imply more equally sized cities. On the other hand, smaller Pareto exponents mean less equal city sizes.

Although the above estimator is a popular means of estimating the Pareto exponent, as Gabaix and Ioannides (2004) showed, this approach provides biased estimates in small samples. Gabaix and Ibragimov (2011) derived a simple and practical solution to correct this bias. They showed that a shift of 1/2 for the rank is optimal and could cancel the bias up to a leading order. Therefore, it is better to estimate equation (3) instead of equation (2):

$$\log(y - 1/2) = \log A - \alpha' \log x. \tag{3}$$

They further showed that the standard error on the Pareto exponent is not the OLS log-log rank-size regression standard error, but is asymptotically $\hat{\alpha}'\sqrt{2/n}$, where n is the corresponding sample size. The OLS log-log rank-size regression for equation ⑵ considerably underestimates the true standard error of the Pareto exponent, too often leading to the rejection of the true numerical value of the Pareto exponent.

The second purpose of this study is to explain the variations in the Pareto exponent over time. We control for the variables that could influence the size distribution of cities over time, including transport costs, land rent, scale economies, and level of development, etc. The empirical model in our estimation will be:

$$\alpha_t = f(URDEN_t, PCAR_t, GOV_t, 2ND/3RD_t, PGDP_t, u_t), \tag{4}$$

where t is the index for the year; *URDEN* is the population density of urban area; *PCAR* is the per capita passenger cars; *GOV* is the percentage of the local government's expenditure in total government expenditure; 2*ND*/3*RD* is the ratio of manufacturing output relative to services output; and *PGDP* is the index of per capita gross domestic product calculated at constant prices and the year 1978 = 100. The dependent variable is the Pareto exponent obtained from equations ⑵ and ⑶.

As more and more economic activities are concentrated in cities, firms and households start to suffer from higher land rents and more congestion. This provides an incentive for firms and households to relocate from the congested centers to the relatively uncluttered periphery. Thus, we use the variable *URDEN* to capture the land rent. Higher population density of an urban area means higher land rent. It is therefore expected that higher land rents will promote population dispersal and the estimated coefficient of *URDEN* will be positive.

Furthermore, since firms want to earn maximal profits and the higher transport costs imply higher distribution costs for firms, firms will as a result be motivated to move away from core cities to peripheral areas where demand is relatively high to reduce distribution costs. For households the measure of transport costs is similar to the degree of convenient transportation which thereby indicates the quality of urban life. As mentioned above, the transport cost is captured by *PCAR*. Fewer per capita passenger cars means a higher transport cost. We can therefore expect a higher transport cost to cause population dispersal, and the coefficient of *PCAR* will be negative.

Several empirical studies have attempted to examine the determinants of city growth (Glaeser et al., 1995; Anderson and Ge, 2004). Anderson and Ge (2004) stated that in the early stage of industrialization, manufacturing cities tend to expand more quickly than service-oriented cities. As the economy industrialized, its structure begins to shift from a manufacturing to a service sector economic base. The service sector will create a great number of job opportunities and absorb a large number of agricultural sector and manufacturing sector laborers. The service sector may be more skill-intensive than the manufacturing industries and thus grow faster. After industrialization, the service-oriented cities tend to expand more quickly than manufacturing cities. Thus, Anderson and Ge (2004) used the ratio of manufacturing output relative to service output to investigate how the manufacturing sector and service sector affect urban growth. They found that the population growth is negatively related to the ratio of manufacturing to service output, implying that expansion of the service sector brings higher growth opportunities and attracts more migrants.

Based on the past related literature, which indicates that industrial structure in an economy may play an important role in urban growth, we use the ratio of manufacturing to service output (2*ND*/3*RD*) to investigate the importance of the manufacturing sector relative to the service sector for population dispersal or concentration.

Government expenditures can affect the quality of urban life, as higher government expenditures imply better public amenities and health care systems. They give people incentives to relocate to large cities to enjoy a better quality of urban life. This study adopts the local government expenditure in total government expenditure[2] to measure the role and impact of governments on the city-size distribution. The role and impact of government expenditures on the city-size distribution is represented by *GOV* and its squared term *GOV*2, as the effect of *GOV* on population concentration is non-linear. We can therefore expect that higher government expenditures will promote population concentration and the estimated coefficients on *GOV* will be negative, and the estimates of *GOV*2 will be positive.

In initial phases of development, economic growth promotes population concentration. Core cities usually have sufficient resources, so production and population tend to relocate to those core cities experiencing the highest economic development. During phases of low levels of development, the population will become more concentrated. In phases with high levels of development, the benefit to concentration at core cities begins to decline. Thus production and population tend to

2 Because the total government expenditures include the expenditure from the central government, which has a wide variety of national expenditure, such as national defense, the administrative expenses, and various operating expenses at the level of central government, we exclude the expenditures from the central government to focus more on local issues. The expenditure of the local governments focuses more on improving the quality of urban life, which is closer to our intention in this study.

relocate to peripheral cities to obtain more benefits. Thus, the high development phases promote population dispersal. As mentioned above, the level of development is measured by *PGDP*. We can therefore expect the estimated coefficient for *PGDP* to be positive.

As noted earlier, we first estimate the Pareto exponent to measure the size distribution of cities for each year. Second, we explain the variations in the Pareto exponent over time. In the first stage, we use a data set for the Chinese city population from 1984 to 2008 obtained from the *China City Statistical Yearbook*, *China Population Statistical Yearbook*, and *China Population and Employment Statistical Yearbook* (NBS, 1985-2003; 2004-2006; 2008-2009).

There are three different administrative levels of cities in the Chinese urban system: county-level cities, prefecture-level cities and central municipalities. For cities at the prefecture-level and above, information on both "Shiqu" (urban areas) and "Diqu" (urban areas and rural areas) are reported. Regarding city population, there are four categories in the *China City Statistical Yearbook* (NBS, 1985-2003) including: (1) total population in Shiqu; (2) total population in Diqu; (3) non-agricultural[3] population in Shiqu; and (4) non-agricultural population in Diqu. Since the data for the third category is the most complete, this paper uses data on the non-agricultural population in urban areas (Shiqu) to measure the city size. Data pertaining to the year 2006 are limited to the *China Population and Employment Statistical Yearbook* (NBS, 2008-2009), which does not report data on non-agricultural population in county-level cities, and thus we omit data for this year.

Regarding the sample selection, Rosen and Resnick (1980) suggested two possible criteria to choose the cut-off for the sample of cities: (1) a fixed number of cities; and (2) a fixed population size threshold. Wheaton and Shishido (1981) also suggested a third criterion where the sample only accounts for some given proportion of the country's population. However, it is easy to see that the other two criteria are problematic: the first being because the number of urban cities in China has increased dramatically from 193 in 1978 to 655 in 2008. Thus, using the first criterion to explore the size distribution of cities is inappropriate. Second, the total population in China has increased from 962.59 million in 1978 to 1,328.02 million in 2008. Hence using the third criterion to explore the size distribution of cities will cause the sample size to be quite different in each year. Thus the third criterion is also inappropriate. In our study we use the second criterion to explore the size distribution of cities. That the number of cities increases over time is an

3 The non-agricultural population is those engaged in non-agricultural vocations and children brought up by non-agricultural workers.

important feature of the Chinese urban system.

The threshold in our study is set at average size of Chinese cities for each year (as shown in column 3 of Table 3). Table 3 provides a background summary of the changes in the number and average size of Chinese cities from 1984 to 2008.

Data for the second stage regression, which seeks to uncover the factors that influence the Pareto exponent, *PCAR*, *GOV*, 2*ND*/3*RD* and *PGDP* are obtained from the *China Statistical Yearbook* (NBS, 1985–2009), and *URDEN* is obtained from the *China City Statistical Yearbook*, *China Population Statistical Yearbook*, and *China Population and Employment Statistical Yearbook* (NBS, 1985–2003;

Table 3. Summary statistics of Chinese city size (10,000 people)

Year	Number of cities	Mean city size	City size standard deviation	Minimum city size	Maximum city size
1984	295	37.3312	66.2299	0.67	672.57
1985	324	36.4970	65.0557	0.25	687.13
1986	348	35.0398	64.3861	0.04	698.73
1987	382	33.9460	63.5186	0.26	711.13
1988	430	32.3319	61.6566	0.29	722.86
1989	449	32.5383	61.9405	0.44	743.45
1990	466	32.2508	61.7009	0.31	749.65
1991	477	32.3167	61.6548	0.35	752.82
1992	517	31.7980	60.0030	0.41	756.1
1993	570	31.0692	59.3067	0.59	810.35
1994	622	30.8124	58.0956	0.47	824.91
1995	640	31.2747	58.5645	0.51	833.8
1996	665	31.2468	58.3826	0.52	841.75
1997	666	32.0926	59.8207	0.54	868.79
1998	664	33.7423	64.3778	0.55	893.72
1999	665	33.5301	63.3881	1.44	923.19
2000	659	34.4962	65.5735	1.48	938.21
2001	657	35.9092	68.2602	1.56	983.84
2002	641	39.0441	72.3850	3.96	1003.08
2003	653	41.5796	77.6635	1.5825	1024.993
2004	655	43.6127	82.5349	1.5773	1079.997
2005	656	45.9547	87.2034	1.5679	1128.371
2007	656	48.1736	92.5890	1.61	1174.053
2008	655	49.0951	94.2752	1.6295	1192.237

2004–2006; 2008–2009). Table 4 lists the definitions of the variables used in the study, their descriptive statistics, and expected signs.

Table 4. Descriptive statistics

Variable	Label	Sing	Mean	Std.
Non-agricultural population in urban area	*X*			
The rank of the city	*Y*			
Total population of urban area divided by urban area (population density of urban area)	*URDEN*	+	268.6364	20.92824
Per capita passenger cars	*PCAR*	−	0.011645	0.009441
The percentage of the local government expenditure in total government expenditure (%)	*GOV*	−	67.7166	6.693158
Square of the expenditure of the local government	*GOV*2	+	4628.47	874.5394
The ratio of manufacturing output relative to services output	2*ND*/3*RD*	−	1.314409	0.158451
Indices of per capita gross domestic product	*PGDP*	−	481.8756	292.1119

IV. Empirical results

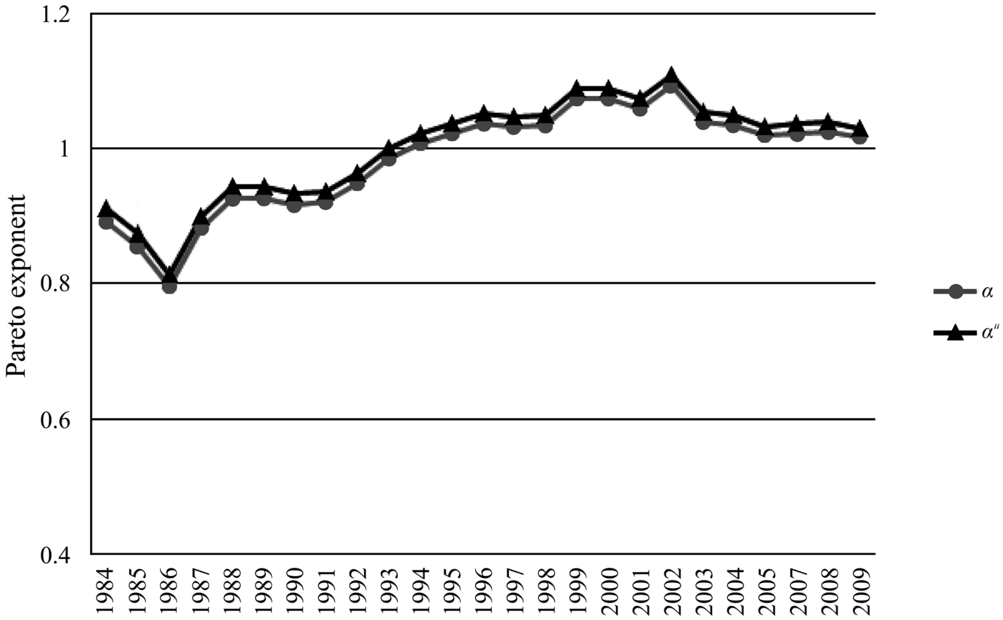
In this section, we first present the empirical results for the Pareto exponent from 1984 to 2008, and then follow the two regressions (i.e., Eq. (2) and Eq. (3)) described in Section 3. Lastly, we seek to explain the variation in the Pareto exponent.

A. Pareto exponent for cities

We calculate the two previously mentioned measures for the entire period from 1984 to 2008, and the results are displayed in Fig. 2. The Pareto exponents for China's cities are decreasing from 1984 to 1986 and then exhibit a slight upturn. This implies that the size distribution of cities is relatively even. After 1994, the value of the Pareto exponent was significantly greater than 1, implying that the size distribution of Chinese cities is more equal than would be predicted by Zipf's law.

In 1980, Chinese authorities announced a development scheme for its urban areas at a national urban planning conference. It focused on how to balance urban and rural development and how to avoid some of the disadvantages of urban development. The urban developmental strategy indeed focused on controlling the large-sized cities and encouraging growth of small-sized cities. The guideline for the

Fig. 2. Evolution of estimations of the Pareto exponent estimated by Eq. (2) and Eq. (3)



scheme was based on the rule of "strict control of the large-sized cities, reasonable development of the medium-sized cities, and aggressive development of the small-sized cities." To investigate the above scenario, we then drew a distinction between large and small cities using a criterion of population size threshold. The threshold in the study is set at the average size of Chinese cities for each year, which means the threshold will shift from year to year. Cities with population greater than the average size of Chinese cities are called large cities, and cities with populations below the average are called small cities.

Fig. 3-1 and 3-2 both plot the Pareto exponent for large cities and small cities from Eq. (2) and Eq. (3), respectively. Unlike in Fig. 2, in Fig. 3-1 the Pareto exponent only for small cities is decreasing from 1984 to 1986 and then exhibits a slight upturn. Higher values imply that the size distribution of cities is becoming more even. After 2002, the value of the Pareto exponent began to decrease again. Overall, the Pareto exponents for small cities are less than 1. This implies that the size distribution of Chinese cities is more uneven than would be predicted by Zipf's law. However, the Pareto exponent for large cities is more stable than that for the full sample (Fig. 2) and small cities (Fig. 3-1). It is greater than 1 and roughly about 1.4. This indicates that large cities are more evenly distributed than would be predicted by Zipf's law. Fig. 3-2 plots the result from Eq. (3); the trend

Fig. 3-1 The Pareto exponents for large cities and small cities from Eq. (2)

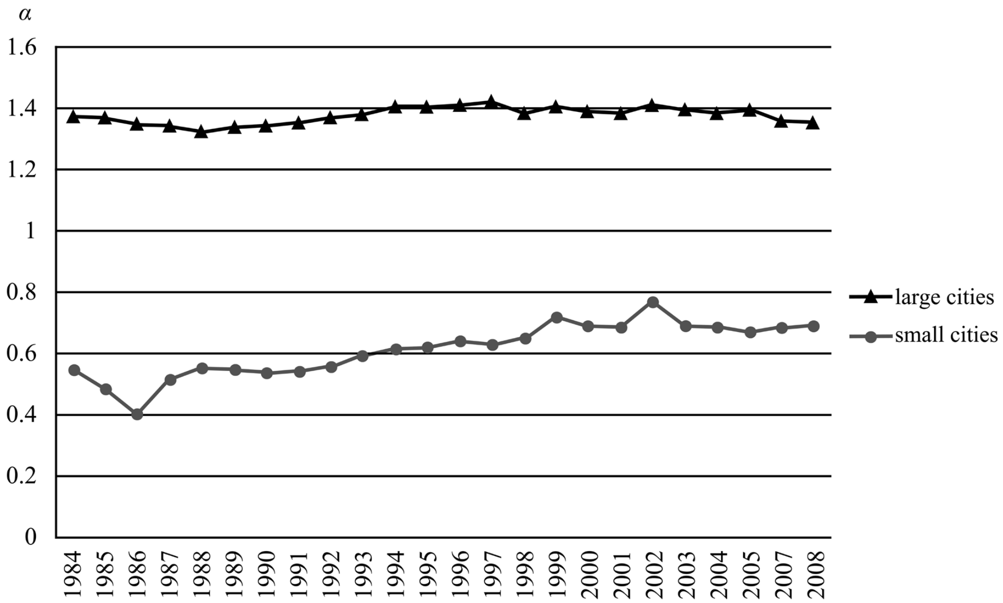
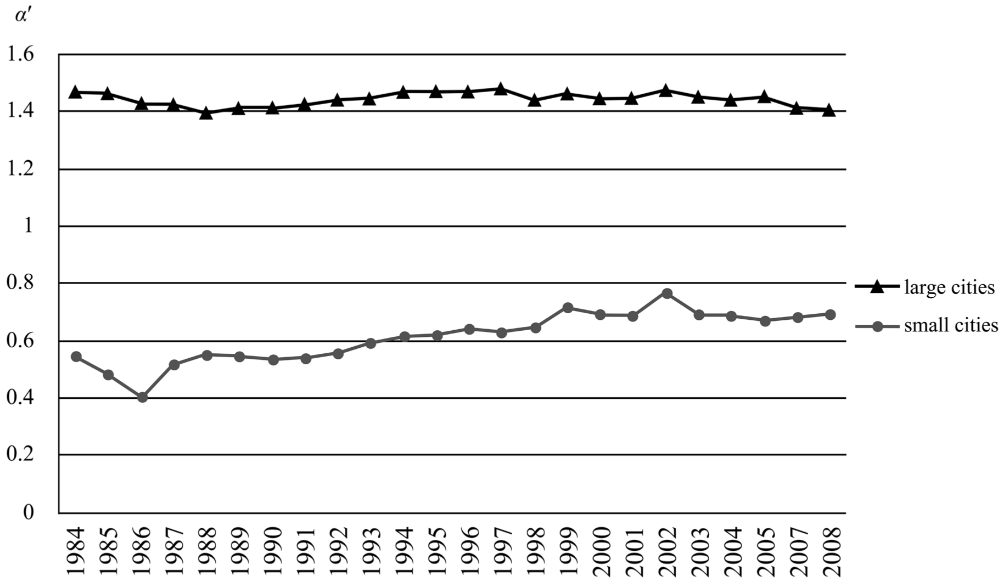


Fig. 3-2 The Pareto exponents for large cities and small cities from Eq. (3)



for Eq. (3) is generally showing a similar pattern.

Table 5 summarizes the results for the Pareto exponent using Eq. (2) and Eq. (3). Even though the two approaches have a similar evolution, there are subtle differences between them. The OLS log-log rank-size regression estimated coefficients, the mean and standard errors of α, are more strongly downward biased in small samples for large cities compared with the sample size comprising all cities with small cities (a detailed result of the Pareto exponent in each year is given in the Appendix).

Table 5. Descriptive statistics of Pareto exponents

	All cities		Large cities		Small cities	
	α	α'	α	α'	α	α'
Mean	0.984319	0.999963	1.376388	1.443987	0.613659	0.614719
Std. Dev.	0.078382	0.077187	0.027292	0.024391	0.085946	0.085983
Min.	0.795526	0.812142	1.321747	1.395172	0.406435	0.406435
Max.	1.093388	1.109107	1.420635	1.481374	0.767428	0.768621
Obs.	24	24	24	24	24	24

Note: α is the estimated Pareto exponent using Eq. (2) and α' is the estimated Pareto exponent using Eq. (3).

In this study discussed above, we can see that the mean of the Pareto exponent for large cities is larger than that for small cities, and the mean of all cities lies between that for large cities and small cities. Overall, Chinese cities have become more evenly distributed than before. Besides, these findings indicate that the size distribution of cities is more even for large cities than for smaller cities. In the next section we will investigate what caused the large cities to be more evenly distributed than the smaller ones.

B. Explaining variation in the Pareto exponent

As discussed above, we find that large cities are more evenly distributed than the smaller ones, and thus a secondary purpose of this study is to explore the determinants of city-size distribution based on Eq. (4). In anticipating that the factors determining the city-size distribution of large cities would be different from those of small cities, we fitted the model differently to the large cities from small cities.

Table 6 presents the empirical results using Eq. (2) and (3) to estimate the

Table 6. Regression results of population distribution model

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Large cities				Small cities			
Dependent variable	α	α'	α	α'	α	α'	α	α'
Explanatory variables								
URDEN	−0.000055	−0.000128			0.000238	0.000237		
	(0.000189)	(0.000189)			(0.000381)	(0.000381)		
PCAR	−16.857050***	−15.231550***	−19.044130***	−18.033340***	−25.754620***	−25.746820***	−29.245610***	−29.282470***
	(2.760497)	(2.520503)	(2.524824)	(2.655950)	(5.490842)	(5.494913)	(6.892877)	(6.934616)
GOV			−0.047490***	−0.056820***			−0.009483	−0.009810
			(0.009027)	(0.009925)			(0.033089)	(0.033285)
*GOV*2			0.000383***	0.000446***			0.000082	0.000084
			(0.000072)	(0.000080)			(0.000258)	(0.000260)
2*ND*/3*RD*	0.096904***	0.122065***			−0.019817	−0.019518		
	(0.026582)	(0.026970)			(0.074791)	(0.075424)		
PGDP	0.000598***	0.000527***	0.000564***	0.000519***	0.001042***	0.001042***	0.001135***	0.001136***
	(0.000094)	(0.000086)	(0.000078)	(0.000081)	(0.000180)	(0.000181)	(0.000209)	(0.000210)
Constant	1.172164***	1.241280***	2.770720***	3.189016****	0.373584**	0.374479**	0.671529	0.683505
	(0.059375)	(0.061914)	(0.275344)	(0.302524)	(0.142146)	(0.143022)	(1.044609)	(1.050549)
R-squared	0.646778	0.557954	0.745105	0.652776	0.837219	0.835478	0.837456	0.835693
Adjusted R-squared	0.572416	0.464892	0.691442	0.579676	0.802949	0.800842	0.803236	0.801102
Observations	24	24	24	24	24	24	24	24

Note: Figures in parentheses are associated *t* values. *significant at 10% level; **significant at 5% level; ***significant at 1% level.

Pareto exponent as the dependent variable. Columns (1), (2), (3) and (4) report the results for large cities. Columns (5), (6), (7) and (8) report the results for small cities. The factors determining the city-size distribution of large cities are quite different from those for small cities. For large cities the five variables *PCAR*, *GOV*, *GOV*2, 2*ND*/3*RD* and *PGDP* yield significant estimates, but *GOV*, *GOV*2 and 2*ND*/3*RD* turn out to be insignificant in the case of small cities. For small cities, the two variables *PCAR* and *PGDP* yield significant estimates. The estimate for *URDEN* is insignificant in both cases of large cities and small cities.

The variable *URDEN* is used to capture the land rent. The estimate for *URDEN* is insignificant in the case of both large and small cities. This implies that in both cases of large cities and small cities the land rent is not the main factor affecting the city size distribution.

The coefficient of *PCAR* is negative and significant in the case of large cities and small cities indicating that higher per capita passenger cars means lower transport cost, and lower transport cost will promote population concentration.

The role and impact of government on city-size distribution is represented by the pair *GOV* and its squares. The findings indicate that the coefficient of *GOV* is negative, while the coefficient of *GOV*2 is positive, and both are significant in the case of large cities. This suggests that higher government expenditures will promote population concentration in the case of large cities. The government expenditure variable *GOV* and its squares, however, are insignificant in small cities. This implies that in the case of small cities the government expenditure is not the main factor that affects its city size distribution.

We use the ratio of manufacturing output relative to service output to investigate how the manufacturing sector and service sector affect the size distributions of Chinese cities. The result shows that the estimated coefficient of 2*ND/*3*RD* is positive and significant in the case of large cities, implying that expansion of the service sector would promote population concentration. That is, expansion of the service sector will bring higher growth opportunities and attract more migrants in the case of large cities. However, 2*ND/*3*RD* is insignificant in the case of small cities.

The level of development is measured by *PGDP*; the result indicates that *PGDP* is highly significant and positive in the case of large cities and small cities. Accordingly, our findings provide support that in initial phases of development, economic growth promotes population concentration. Since core cities usually have sufficient resources, production and population tend to relocate to those core cities experiencing the highest economic development, thus causing the population to become more concentrated.

V. Conclusions

In the urban economics literature, a number of studies have emerged seeking to explain how and why the size distribution of cities differs across countries. While many studies in the literature analyze the size distribution of cities for single countries and in different periods, few investigate the size distribution of cities with time series data for a single country. In this paper, we have attempted to examine the size distribution of Chinese cities over the period from 1984 to 2008. We have also formulated and estimated a model that is designed to explain the processes and factors standing behind and shaping the size distribution of cities.

Two of these findings can be briefly summarized as follows. First, we present the empirical results regarding the evolution of the Pareto exponent for China from 1984 to 2008. Furthermore, we find that the size distribution of large cities is more stable and roughly about 1.4. However, compared to large cities, the size distribution of small cities experiences some fluctuation, and is less than 1 and roughly about 0.7. The results therefore indicate that the size distribution of cities is more even for large cities than for smaller cities. Overall the Chinese cities have become more evenly distributed than before. These findings agree with those of Song and Zhang (2002), who observed that the Chinese cities had become more evenly distributed. The present study enhances Song and Zhang's (2002) findings by providing a much more detailed examination of Chinese cities. We explain the processes and factors that lie behind and shape the distribution of city size. Furthermore, we investigate what caused the large cities to be more evenly distributed than smaller cities.

Second, as discussed above, we find that both large cities and small cities have two variables, *PCAR* and *PGDP*, that yield significant estimates. The coefficient of *PCAR* is negative and significant. This implies that lower transport cost will promote population concentration and hence city size will be more unevenly distributed. The level of development is measured by *PGDP*. As expected, at lower levels of *PGDP* the relationship will be positive, which means in initial phases of development, economic growth promotes population concentration.

In this paper, we find that the factors determining the city-size distribution of large cities are quite different from those for small cities. The role and impact of government on city-size distribution is represented by the pair *GOV*, and its squares are significant in the case of large cities, but turn out to be insignificant in the case of small cities. Since large cities usually have better public amenities, health care systems and sufficient resources compared to small cities, population tends to relocate to those large cities experiencing the highest quality of urban life.

We also find that industrial structure in an economy play an important role in city size distribution, especially in the case of large cities. Expansion of the service sector would promote population concentration, that is, expansion of the service sector will bring higher growth opportunities and attract more migrants in the case of large cities. The estimate for *URDEN* is insignificant in both cases of large cities and small cities. It implies that in both cases, the land rent is not the main factor affecting the city size distribution. One reason for this result could be that large cities have sufficient resources and better quality of urban life, so they have higher growth opportunities than small cities. Thus, even if firms and households suffer from higher land rents and more congestion, they still prefer to locate in large cities.

In conclusion, transportation costs and the level of development have had a significant effect on the size distribution of cities. A possible explanation for larger cities being more evenly distributed than smaller cities is that the quality of urban life and expansion of the service sector bring higher growth opportunities and attract more migrants. Thus, government policy and industrial structure may play an important role in finding ways to reduce the sizes of cities.

References

Alperovich, Gershon

1993 "An Explanatory Model of City-size Distribution: Evidence from Cross-country Data," *Urban Studies* 30(9): 1591–1601.

Anderson, Gordon and Ying Ge

2004 "Do Economic Reforms Accelerate Urban Growth? The Case of China," *Urban Studies* 41(11): 2197–2210.

2005 "The Size Distribution of Chinese Cities," *Regional Science and Urban Economics* 35(6): 756–776.

Auerbach, F.

1913 "Das Gesetz der Bevölkerungskoncentration," *Petermanns Geographische Mitteilungen* 59: 74–76. (Auerbach, F., 1913, "The Law of Population Concentration," *Petermann's Geographical Information* 59: 74–76.)

Gabaix, Xavier

1999 "Zipf's Law for Cities: An Explanation," *The Quarterly Journal of Economics* 114(3): 739–767.

Gabaix, Xavier and Rustam Ibragimov

2011 "Rank-1/2: A Simple Way to Improve the OLS Estimation of Tail Exponents," *Journal of Business and Economic Statistics* 29(1): 24–39.

Gabaix, Xavier and Yannis M. Ioannides
2004 “The Evolution of City Size Distributions,” pp. 2341–2378 in J. V. Henderson and J. F. Thisse (eds.), *Handbook of Regional and Urban Economics*, Vol. 4. Amsterdam: North-Holland.
Gan, Li, Dong Li, and Shunfeng Song
2006 “Is the Zipf Law Spurious in Explaining City-size Distributions?” *Economics Letters* 92(2): 256–262.
Giesen, Kristian and Jens Südekum
2011 “Zipf’s Law for Cities in the Regions and the Country,” *Journal of Economic Geography* 11(4): 667–686.
Glaeser, E. L., J. A. Scheinkman, and A. Shleifer
1995 “Economic Growth in a Cross-section of Cities,” *Journal of Monetary Economics* 36(1): 117–143.
Hill, Bruce M.
1975 “A Simple General Approach to Inference about the Tail of a Distribution,” *The Annals of Statistics* 3(5): 1163–1174.
Krugman, Paul
1996 *The Self-organizing Economy.* Oxford: Blackwell Publishers.
NBS (National Bureau of Statistical of China)
1985–2003 *China City Statistical Yearbook*. Beijing: China Statistical Press.
1985–2009 *China Statistical Yearbook*. Beijing: China Statistical Press.
2004–2006 *China Population Statistical Yearbook*. Beijing: China Statistical Press.
2008–2009 *China Population and Employment Statistical Yearbook*. Beijing: China Statistical Press.
2009 *China City Statistical Yearbook*. Beijing: China Statistical Press.
Nitsch, Volker
2005 “Zipf Zipped,” *Journal of Urban Economics* 57(1): 86–100.
Parr, John B.
1985 “A Note on the Size Distribution of Cities over Time,” *Journal of Urban Economics* 18(2): 199–212.
Peng, Guohua
2010 “Zipf’s Law for Chinese Cities: Rolling Sample Regressions,” *Physica A: Statistical Mechanics and Its Applications* 389(18): 3804–3813.
Rosen, Kenneth T. and Mitchel Resnick
1980 “The Size Distribution of Cities: An Examination of the Pareto Law and Primacy,” *Journal of Urban Economics* 8(2): 165–186.
Singer, H. W.
1936 “The ‘Courbe des Populations.’ A Parallel to Pareto’s Law,” *The Economic Journal* 46(182): 254–263.
Song, Shunfeng and Kevin Honglin Zhang
2002 “Urbanisation and City Size Distribution in China,” *Urban Studies* 39(12): 2317–2327.

Soo, Kwok Tong
2005 "Zipf's Law for Cities: A Cross-country Investigation," *Regional Science and Urban Economics* 35(3): 239–263.
2007 "Zipf's Law and Urban Growth in Malaysia," *Urban Studies* 44(1): 1–14.
Wheaton, William C. and Hisanobu Shishido
1981 "Urban Concentration, Agglomeration Economies, and the Level of Economic Development," *Economic Development and Cultural Change* 30(1): 17–30.
Zipf, George Kingsley
1949 *Human Behavior and the Principle of Least Effort.* Cambridge: Addison-Wesley.

Appendix: Results of OLS regression of Eq. (2) and (3)

	(1) All cities						
Year	Cities	α	s.e	α'	s.e	Cities	α
1984	295	0.891483	0.017445	0.912649	0.075146	69	1.371986
1985	324	0.855572	0.019448	0.874545	0.068711	73	1.369760
1986	348	0.795526	0.020239	0.812142	0.061568	83	1.347026
1987	382	0.881673	0.016803	0.899211	0.065065	87	1.341184
1988	430	0.926931	0.015437	0.944195	0.064394	99	1.321747
1989	449	0.925869	0.015322	0.942543	0.062906	105	1.338571
1990	466	0.916601	0.015264	0.932610	0.061097	108	1.341942
1991	477	0.920616	0.015031	0.936411	0.060635	110	1.351779
1992	517	0.947940	0.014798	0.963415	0.059922	120	1.371146
1993	570	0.985468	0.013887	1.000681	0.059275	130	1.380482
1994	622	1.007005	0.013489	1.021640	0.057932	139	1.406499
1995	640	1.022707	0.013286	1.037327	0.057988	145	1.407253
1996	665	1.037625	0.012536	1.052051	0.057695	152	1.410904
1997	666	1.032353	0.012873	1.046682	0.057358	153	1.420635
1998	664	1.033980	0.012245	1.048406	0.057539	149	1.382757
1999	665	1.074813	0.011673	1.089987	0.059776	139	1.403004
2000	659	1.074473	0.011733	1.089766	0.060035	149	1.389344
2001	657	1.058397	0.011619	1.073370	0.059222	146	1.385346
2002	641	1.093388	0.009790	1.109107	0.061953	142	1.410879
2003	653	1.039071	0.010653	1.053509	0.058304	149	1.393545
2004	655	1.035849	0.010696	1.050245	0.058034	149	1.382879
2005	656	1.019075	0.010635	1.033092	0.057043	152	1.393621
2007	656	1.022459	0.010266	1.036574	0.057235	149	1.358233
2008	655	1.024774	0.010307	1.038954	0.057410	145	1.352793

(2) Large cities			(3) Small cities				
s.e	α'	s.e	Cities	α	s.e	α'	s.e
0.018384	1.469795	0.250234	226	0.545552	0.016159	0.547335	0.051489
0.017860	1.463734	0.242279	251	0.485277	0.017809	0.486731	0.043448
0.016706	1.431691	0.222242	265	0.406435	0.017605	0.406435	0.035309
0.016003	1.422709	0.215711	295	0.516414	0.016263	0.517727	0.042629
0.014528	1.395172	0.198301	331	0.554503	0.016486	0.555750	0.043200
0.013464	1.410516	0.194670	344	0.546280	0.015922	0.547447	0.041742
0.013191	1.412789	0.192256	358	0.537051	0.015355	0.538159	0.040224
0.012951	1.422389	0.191795	367	0.542871	0.015058	0.543968	0.040156
0.011349	1.439019	0.185777	397	0.556253	0.015165	0.557284	0.039555
0.009522	1.445909	0.179343	440	0.593891	0.014547	0.594899	0.040108
0.008477	1.470606	0.176402	483	0.612767	0.013830	0.613732	0.039493
0.008080	1.469458	0.172579	495	0.621809	0.013888	0.622754	0.039585
0.007819	1.471223	0.168761	513	0.641909	0.013053	0.642849	0.040139
0.007673	1.481374	0.169369	513	0.629366	0.013320	0.630284	0.039354
0.008493	1.442117	0.167079	515	0.649928	0.013336	0.650889	0.040562
0.008348	1.466867	0.175954	526	0.714954	0.012820	0.716046	0.044153
0.007644	1.449642	0.167951	510	0.692238	0.012655	0.693269	0.043414
0.008627	1.445886	0.169228	511	0.687295	0.012532	0.688329	0.043063
0.009552	1.473551	0.174878	499	0.767428	0.010469	0.768621	0.048661
0.010292	1.452539	0.168287	504	0.691910	0.011392	0.692951	0.043652
0.009492	1.441455	0.167002	506	0.686553	0.011391	0.687582	0.043228
0.009188	1.451954	0.166550	504	0.670345	0.010882	0.671343	0.042291
0.010167	1.415038	0.163942	507	0.685314	0.010949	0.686345	0.043108
0.010691	1.410263	0.165627	510	0.691476	0.010928	0.692530	0.043368

中國城市人口分布與 Zipf 法則

林俊宏
淡江大學產業經濟學系教授

鄭美君
臺灣經濟研究院副研究員

楊志海
國立中央大學經濟學系教授

摘　　要

中國自 1978 年實施經濟改革後，經濟便開始快速成長，然而經濟改革不只爲中國經濟帶來快速的發展，亦帶來了都市化現象。由於經濟改革造成快速的都市化，因此中國的都市規模分配之演進值得我們深入探討。本研究將透過中國 1984 年至 2008 年的都市人口探討其都市規模分配的變化。研究結果顯示中國都市的規模分配較 Zipf 法則下更爲平均。本研究亦發現大都市的都市規模分配較小都市的都市規模分配更爲平均，而造成此現象的主要因素爲政府支出及產業結構的改變。本研究亦發現，隨著經濟的發展及較低的運輸成本，促使人口變得較爲集中。

關鍵字：都市分配、Zipf 法則、Pareto 指數、中國