The Size Distribution of Chinese Cities

Gordon Anderson and Ying Ge*

Abstract

This paper uses Chinese urban data to investigate two important issues regarding city size distributions. The nature in which cities of different sizes grow relative to each other is examined and, contrary to the common empirical finding that the relative size and rank of cities remains stable over time, it is found that the period of economic reforms since 1980 was one of significant structural change in the Chinese urban system. The city size distribution remains stable before the reforms but exhibits a convergent growth pattern in the post-reform period. In addition Pearson goodness-of-fit tests are employed to examine directly which theoretical distribution provides the best approximation to the empirical city size distribution. A parallel study of city size distributions in China and U.S. reveals substantial differences with lognormal distributions being the preferred specification in the case of China and Pareto distributions being preferred in the case of the US.

JEL Classification: C24, R12

Keywords: Zipf's Law, Pareto distribution, Convergence

* Correspondence: Gordon Anderson, Department of Economics, University of Toronto, 150 St. George Street, Toronto, Ontario M5S 3G7, Canada. E-mail: anderson@chass.utoronto.ca, Phone: 416-978-4620, Fax: 416-978-6713.

1. Introduction

Two questions have been of long-standing interest to urban economists: the first one is how cities of different sizes grow relative to each other. Most empirical studies suggest that relative size and rank of cities remain stable over time. The results of Eaton and Eckstein (1997) favored a parallel growth pattern over divergent or convergent growth patterns for French and Japanese cities in the period of 1876 - 1990. Black and Henderson (1999) constructed a theoretical model to explain the parallel growth rates of cities. They used data on U.S. cities from 1990 to 1950 to support this pattern of growth. Overman and Ioannides (2001) developed a stochastic transition kernel for the evolution of the distribution of U.S. metropolitan area populations from 1900 to 1990. Their results predominantly suggest persistence in the city size distribution. Dobkins and Ioannides (2000) studied the dynamic evolution of the city size distribution in U.S. and confirmed the city growth in U.S. is relative stable over time. Sharma (2003) examined the growth of Indian city sizes from 1901-1991 and found that the growth of cities may be parallel in the long-run but with short-run deviations from the long-term pattern.

The second question is which theoretical distribution provides the best approximation for the empirical city size distribution. The dominating view in literature is that the Pareto distribution fits best. The early idea that the size distribution of cities can be approximated by a Pareto distribution was mentioned by Auerbach (1913), Zipf (1949) claimed that not only the size distribution of cities follow a Pareto distribution, but also the distribution has a shape parameter equal to 1. In empirical studies of the city size distribution, when cities are ordered by population size, regressing the logarithm of their rank on the logarithm of their population has resulted in a slope coefficient close to minus one in so many instances that the phenomena has acquired the status of the eponymous Zipf's Law[2]. Taking exponents, the relationship can be seen to be a special case of a power rule relating the size rank of a city to some power of its population size rendering the statistical distribution appropriate for the relationship a member of the family due to Pareto (1897) more commonly employed in modeling income distributions. More generally, the Pareto exponents generated from the rank-size regression are not necessarily equal to unity, and this so-called "rank-size rule"is believed to be applicable to almost all countries around world. Rosen and Resnick (1980) estimated the Pareto exponent for 44 countries. Their estimates ranged from 0.81 to 1.96, with a sample mean of 1.14. Soo (2002) investigated a new data set of 75 countries and found the average value of Pareto exponent is 1.1, which is significantly greater than 1 predicted by Zipf's Law.

The conventional rank-size regression has also been extended to test the validity of Pareto distribution in representing the city size distribution. Hsing (1990) argued that the conventional methodology pursued in the literature suffers from the fact that the form of distribution is presumed a priori rather than having been inferred from a general functional form capable of accommodating the Pareto distribution as a special case. He suggested the Box-Cox general transformation function as a more suitable test and found that the Pareto distribution is significantly rejected by U.S. data. In a similar vein Alperovich and Deutsch (1995) also found the Pareto distribution is rejected in favor of a general Box-Cox function when using international data. Ioannides and Overman (2003) used a nonparametric procedure to estimate Gibrat's Law for city growth processes and their results favored Zip's Law.

The rank-size rule emerged from a regularly observed feature of data and lacked theoretical foundation. Gabaix (1999) provides a theoretical basis by entertaining a model of the city size

---

[2]See Gabaix (1999) for citations and an excellent discussion of the literature.

distribution that is the consequence of a stochastic random growth process common to individual city sizes, effectively an application of "la loi de l'effet proportionnel" employed in other contexts by Gibrat (1930, 1931). Initial city size is assumed to suffer a sequence of mutually independent proportionate shocks (in effect a Geometric Brownian Motion) each independent of initial city size and none of which differ substantially from one. Under such conditions Gibrat demonstrates that, after sufficient time, a log-normal size distribution would emerge with a variance that increased over time and, if the process had sufficient positive drift or growth, with a mean that increased over time. Subsequently, by sacrificing the initial size independence assumption, Kalecki (1945) modified Gibrat's contribution to admit log-normality with a non-increasing variance. The distinction between the two is important because the former predicts a distribution that advances in an unbounded fashion whereas the latter predicts a distribution whose location trends through time with the growth rate but whose variance is bounded. Obviously neither the Gibrat nor Kalecki models are consistent with the Pareto distribution[3]. Gabaix establishes the link by positing that the Geometric Brownian Motion for city size is subject to a reflective lower bound. He demonstrates that, provided the growth rate of the number of cities does not exceed the growth rate of their populations, the rank size distribution with a coefficient of one emerges as a consequence. Without the lower bound, city size would follow Gibrat's Law of Proportionate Effects. In a similar fashion Reed (2000, 2003) arrives at a "Double Pareto" power rule by positing the unbounded Geometric Brownian Motion to have progressed for a random period of length T described by an exponential distribution (cities have existed for a random period of time described by the exponential

[3] Early models by Steindl (1965) and Simon (1955), based on assumptions regarding the relationship between the rate at which cities emerged and the rate at which they grew, ran into difficulties because their preconditions are basically counterfactual. Brakman et. al. (1999) by introducing suitably calibrated negative externalities into a general equilibrium location model are able to simulate Rank Size distributions.

distribution). In this case city size follows a power law in both tails. Again for fixed T, city size would follow Gibrat's law.

This paper exploits Chinese data to investigate these two general issues regarding the city size distribution. There are three reasons why the experience of China is of interest. First, most empirical studies in this topic focus on the developed countries and, as a developing country with the largest population and a rapid urbanization process, China provides a "non-developed" comparison case. Secondly, over the last half century, China has experienced substantial economic reform including a transition toward a market economy and integration into the world economy. Whether such a fundamental reform induced a change in the city size distribution is of interest. Thirdly, China has experienced very rapid urbanization process since 1980, with the development of a large number of new cities. This stage is significantly different from the stable city growth stage in most literatures, where the main channel of urban growth is the expansion of existing cities rather than the birth of new ones. A critical element of the Gabaix (1999) argument is that the rate of growth in the number of cities does not exceed the population growth rate and it is instructive to observe the consequences for his theory when this condition is not met.

This paper makes two contributions. First, in examining the evolution of the Chinese urban system over time it finds that economic reforms engendered a structural change in the city size distribution. In the pre-reform period the city size distribution is relatively stable while in post-reform period there is a significant convergence trend in city growth. Secondly, alternative functional forms suggested by theory were examined to see which provided the best approximation to the distribution of city sizes. Rather than relying on the value of a regression coefficient as evidence of a distributional form, maximum likelihood estimates of the parameters of theoretically relevant distributions were developed and their congruence with the data was examined directly via Pearson Goodness of Fit tests. For comparison a parallel exercise is undertaken for the USA as a representative of the many countries for which evidence of the power law abounds[4]. The results, whilst supporting the rank size distribution with a coefficient close to one for the USA, strongly reject various power law specifications (and consequently of Zipf's law) for China. In the case of China Gibrat's law appears to describe the situation well prior to the economic reform and Kalecki's reformulation appears to be appropriate in the post reform period.

The paper is organized as follows: Section 2 provides some background to the Chinese city data utilized in the study. Section 3 tests the structural break in the evolution process of Chinese urban system. Details of the various distributions are outlined and examined for thier empirical coherence in section 4. Conclusions are drawn in section 5.

2. City Definition and Data Description

The definitions of a city differ across countries. Chinese cities are officially defined as "urban places"which correspond to local administrative entities. There are three different administrative levels of Chinese cities: municipalities (province-level cities), prefecture-level cities and county-level cities, each having the same status as the province, prefecture and county, respectively. Urban places with townships or lower jurisdictional levels are not treated as cities. This administrative definition of a city is less ideal than the metropolitan areas generally employed in studying city size distributions. Due to data limitations, the population of metropolitan areas in China are not available.

The choice of the lower population threshold is important in the study of city size

[4] See for instance Dobkins and Ioannides (2000) for the USA and Eaton and Eckstein (1997) for France and Japan.

distributions. The Chinese government announces certain criteria distinguishing urban and rural places, cities and towns. In particular there is a lower population boundary for cities which the government altered in 1955, 1963 and 1986. To assemble the time consistent data, 100,000 was adopted as the lower population boundary according to the official criteria announced in 1963 – it defines a city as an urban agglomeration with a total urban population larger than 100,000[5]. The year when the People's Republic of China was established, 1949, was chosen as the starting point of the sample. Economic reforms began at the end of 1970's and accelerated the urbanization progress in China. Roughly speaking 10-year intervals were chosen for the pre-reform period (1949-1980) and 5-year intervals for reform period (1980-1999).

The Chinese urban system data set was compiled from two sources: the urban population of cities from 1949 to 1989 was drawn from *The Forty Years of Urban Development* (State Statistical Bureau, 1990) and the data for 1990 to 1999 was drawn from various issues of *Chinese Urban Statistical Yearbooks* (State Statistical Bureau, 1991-2000).

[Table 1 about here]

Panel A of Table 1 summarizes the change in the number and average size of Chinese cities from 1949 to 1999. There appears to be three stages of city growth. The first is a stable growth stage from 1949 to 1961; the second is a stagnation stage, from 1961 to 1978; and the third is a rapid expansion stage coincident with the economic reforms which started in 1978. A striking feature of the data is the large number of new cities, especially in the reform period. In the pre-reform period the number of cities annually increases by about 5.5%, in the reform period the growth rate increases to 11.4%. In large part urbanization in China takes the form of the creation of new cities,

[5] In China certain urban places with less than 100,000 inhabits are also treated as cities because of their political or economic importance.

a quite different experience from that of developed countries, where urbanization mainly comes from the expansion of existing cities[6] . Another important pattern worthy of note is that the standard deviation of city size decreases after 1980 with average size growing steadily over time. This implies a city size convergence trend not unlike σ-convergence observed in economic growth and income distribution analysis. The last two rows of Panel A compares the number of total cities and the number of cities in our sample. Except the early period, 1949 and 1961, there are about 10-15 cities excluded because the urban population is less than 100,000 inhabitants reflecting the view that the low population boundary of 100,000 inhabitants is reasonable in the case of China.

For comparison data for US city populations was drawn from the "County and City Data books" of the US Census Bureau which is summarized in Panel B of Table 1. Although the metropolitan area population is most commonly used in U.S. studies, for comparability reasons the data for the administratively defined cities (instead of metropolitan areas) is used and a lower population boundary of 100,000 is adopted. In this setup, ensuring differences between city size distributions in the two countries will not due to different city definitions or lower population boundaries. Since the total population in China is much larger than that of the U.S., the uniform low population boundary will exclude a higher proportion of U.S. cities from the sample, a common limitation of cross country studies. However, as will be demonstrated later, redefining the lower bound of a Pareto distribution will not affect its parametric structure.

3. City Size Growth in China

The evolution of city size distributions depends upon two separate processes: the expansion or shrinkage of existing cities and the city birth process. Most studies of developed countries focus

---

[6] For example, Eaton and Eckstein (1998) found "there are no new cities" in France and Japan.

on the growth of existing cities because of the extremely low birth rates of new cities. In the rapid urbanization process of China, new cities play an important role in shaping the city size distribution. Here a convergence hypothesis regarding the relative growth of existing cities is examined, the conventional rank size regression is employed to test the stability of Pareto exponent over time for all cities and finally a Markov chain method is exploited to examine the intra-distributional mobility of all cities, especially the new ones.

To explore the growth pattern of existing cities, a panel data set[7] which excludes all new cities between 1961 and 1999 is constructed. The following cross section regression, familiar in the empirical growth literature, is used to test the β convergence hypothesis of city size growth:

$$\frac{1}{T}\log\left(\frac{y_{i,t_0+T}}{y_{i,t_0}}\right) = a - \left(\frac{1-e^{-\beta T}}{T}\right)\log(y_{i,t_0}) + \varepsilon_{it_0,t_0+T} \qquad [1]$$

Here y_{it} is the size of city i at time t, a is constrained to be identical across cities, and coefficient β is interpreted as the speed of convergence. The nonlinear least-squares results are reported in Table 2.

[Table 2 about here]

As Table 2 reveals there is significant β convergence in both pre-reform and reform periods for existing cities. Moreover, the convergence significantly speeds up in the reform period. This pattern is clearly shown in Figure 1, where the annual growth rate and the initial city size are negatively correlated.

Returning to the full sample, the conventional rank size regression is employed to examine the evolution of the Pareto exponent over time. 1980, the first observation in the reform years, is chosen to be the benchmark. Time dummies and interaction terms are included to capture the different intercepts and slopes in each year from 1980. The result is reported in Table 3.

---

[7] This data set includes 157 cities and only covers the period from 1961 to 1999 because the paucity of observations in 1949 render the balanced panel data set too small.

[Table 3 about here]

As may be observed, in 1980 there is a significant structural break in the evolution of the Pareto exponent θ. For the pre-reform period, 1949-1980, the estimated θ is not statistically significantly different from -1 at the 1% level. From 1980 to 1999, the size distribution of the cities is increasingly convergent except for a slight reversal from 1994 to 1999. The Pareto exponents are significantly higher than 1 in the reform period, which implies that the distribution of city size has become more equal than Zipf's law would predict. The intercepts are subject to more variation since they are directly dependent on the sample sizes in each year. This pattern is different from the parallel growth pattern of developed countries (Eaton and Eckstein, 1997, Dobkins and Ioannides, 2000).

A Markov chain method is often used to analyze intra-distribution income dynamics (Quah, 1993). Here Dobkins and Ioannides (2000) is followed in assuming the size distribution of exiting cities follows a first-order homogeneous Markov process. Panels A and B of Table 4 report the average decade transition matrix in the pre-reform and reform periods respectively. The chosen intervals were based on ¼, ½, ¾,1 and 2 times the sample mean.

[Table 4 are about here]

Two features stand out. First, the diagonal entries indicate that larger cities have higher persistence and smaller cities are more likely to move to upper categories. The trend of upward concentration is more significant in the pre-reform period than in the reform period which could be due to new entries in the lower tail. To examine this, the balanced panel of 157 cities existing from 1961 to 1999 is employed to calculate the distribution abstracted from new entries and exits.

The results, not reported here[8], indicate that a significant concentration toward the upper tail of the distribution remains, suggesting real growth in city size as well as new entrants is causing the phenomenon. This pattern is quite different from the case of France and Japan (Eaton and Eckstein, 1998).

A second feature is apparent in the last row of the table which shows the frequency distribution of new entrants. The relative size of new cities is significantly different before and after 1980. In the pre-reform period, 77% of new cities are smaller than half of the mean. In the reform period, 78% of new cities are medium size (from half the mean to twice the mean). Recalling that the observation interval in the reform period is 5 years or shorter this is truly remarkable. It implies that settlements too small to be classified as cities in the distribution at the onset of the interval between observation periods have grown to such a size as to merit their entrance in the middle of the distribution within 5 years.

Combinations of the growth processes of existing cities and the pattern of new entrants explain the difference in city size evolution in pre-reform and post-reform periods. In the pre-reform period, the new entrants concentrate in the lower tail and raise the relative size ranking of existing cities. For existing cities, small cities expand into the middle range with the big cities in the upper tail exhibiting strong downward immobility. The combination of these two movements results in a relatively stable size distribution which is not significantly different from the prediction of Zipf's law. In the reform period, the existing small cities grow faster than larger cities and new entrants concentrate around the middle sizes. As a result, the size distribution exhibits a strong convergence trend.

[8] Available from the authors on request.

4. Statistical Tests of Alternative Distributions

Here the alternative functional forms used to describe the city size distribution are outlined and the Pearson Goodness-fit test is employed to examine which distribution provides the best approximation to the empirical city size distribution in China and U.S.

4.1 The alternative distribution representing city size distribution

The rank-size rule for a city size can be derived by transforming the counter cumulative density of the Pareto distribution where city population size is a random variable P with realization p such that: $Prob(P>p) = 1-F(P\leq p) = A^{*}p^{-\theta}$. For a sample of n cities with sizes p_i, i=1,..,n, define the empirical density F^e(p) as $F^{e}(p,p_i) = \frac{1}{n}\sum_{i=1}^{n} I(p_i \leq p)$. The rank size rule emerges from the relationship:

$$Rank(p_k) = n(1-F^{e}(p_k,p_i)) \approx nA^{*}p_k^{-\theta} \qquad [2]$$

This relationship suggests a regression equation for the logarithm of the city population rank of the form:

$$y_i = lnRank(p_i) = lnA - \theta \ln p_i + e_i \qquad [3]$$

where $A=nA^{*}$ and e_i is assumed to exhibit all the statistical properties of an unobserved error process that render [2] a valid regression equation. A special case of this distribution where $\theta = 1$ corresponds to Zipf's law (Zipf [29]).

For $F(P\leq p)$ to be the density function of a Pareto distribution, $A^{*} = p_{min}^{\theta}$ and $p > p_{min} > 0$ which yields a pdf of the form $f(p,p_{min}) = \theta p_{min}^{\theta} p^{-(1+\theta)}$. In the current context p_{min} corresponds to the minimum possible size of a city, the lower reflective boundary in Gabaix's model. Its precise definition has been a matter of theoretical debate in the literature and set by arbitrary fiat in practice. However it is interesting to note that, from the point of view of evaluating empirical support for rank size distributions, providing a lower bound is set somewhere above the true lower bound, the viability of a Pareto distribution is unaffected. Suppose settlements were defined as cities only when their population exceeded $p_{def} > p_{min}$, so that only data for city sizes greater than or equal to p_{def} were entertained, the conditional p.d.f. $f(p \mid p \geq p_{def})$ is readily shown to be $f(p, p_{def}) = \theta p_{def}^{\theta} p^{-(\theta+1)}$, which is exactly the same pdf with a redefined lower bound.

If the lower bound were known[9], [3] could be rewritten as:

$$y^*_i = \ln(Rank(p_i)/n) = -\theta \ln(\frac{p_i}{p_{min}}) + e_i \qquad [4]$$

This specification provides a more efficient estimator of θ given by:

$$\theta_{OLS} = -\frac{\sum_{i=1} y^*_i \ln(\frac{p_i}{p_{min}})}{\sum_{i=1}^{n} (\ln(\frac{p_i}{n}))^2}$$

whereas the maximum likelihood estimator is given by:

$$\theta_{ML} = \frac{n}{\sum_{i=1}^{n} (\ln(\frac{p_i}{p_{min}}))}$$

Under an assumption that cities have existed for a period T where T is governed by an exponential distribution, Reed (2000) develops a Double Pareto distribution for city size where:

$$f(p) = \frac{\alpha\beta}{\alpha+\beta}(\frac{p}{p^*})^{\beta-1} \quad \text{for } p<p^* \qquad [5]$$

$$f(p) = \frac{\alpha\beta}{\alpha+\beta}(\frac{p}{p^*})^{-\alpha-1} \quad \text{otherwise}$$

---

[9]Note that when p_{min} is unknown its maximum likelihood estimator is the first order statistic of the sample. This is of course a biased estimator with a p.d.f. of $n\theta(p_{min}/p)^{n\theta}/p$ which is rendered unbiased by pre-multiplication by $(\theta n-1)/\theta n$ (in the case of Zipf this reduces to $(n-1)/n$).

This distribution yields a rank order rule for both upper and lower tails of the distribution respectively of the form:

$$y^*_i = \ln(Rank(p_i)/n) = \ln(\frac{\beta}{\alpha+\beta}) - \alpha\ln(\frac{p_i}{p^*}) + e_i \quad for\ p_i \geq p* \qquad [6]$$

$$y^*_i = \ln(Rank(p_i)/n) = \ln(\frac{\alpha}{\alpha+\beta}) + \beta\ln(\frac{p_i}{p^*}) + e_i \quad for\ p_i < p*$$

Maximum likelihood estimates for α, β and p* are readily obtained from [6].

Gibrat's (1999) formulation is based on the premise that the initial city size variate P_0 is subject to a sequence of mutually independent proportionate changes e_i, $i = 1,2,...,t$ so that after the passage of time t, $P_t = P_0(1+e_1)(1+e_2)\ ...(1+e_t)$. Assume $|e_i|$ to be small relative to 1 and let $\ln(1+e_k) = u_k$ where u_k is an i.i.d. process with zero mean and variance σ^2 which is small relative to 1 then:

$$lnP_t = \mu + lnP_{t-1} + u_t \qquad [7]$$

with μ (again small relative to one) corresponding to the incremental drift or growth in city size and u_t corresponding to the increment of a drifting Weiner process which, after sufficient passage of time t, renders the distribution of lnP as $N(lnP_0+(\mu-\sigma/2)t\ ,\ \sigma^2 t)$. In this framework the progress of city size is a random walk with drift, there is no long run "equilibrium" city size that is the consequence of economic and physical forces.

Kalecki (1945) proposed an alternative process which, in the present context, replaces [4] with:

$$lnP_t = \eta + \mu t + (1-\lambda)lnP_{t-1} + u_t \qquad [8]$$

where $1 > \lambda > 0$. Kalecki establishes that, after a sufficient passage of time, the distribution of lnP will be $N((\eta+\mu t)/\lambda, \sigma^2/\lambda^2)$. Again μ may be construed as the incremental growth component but in this case the logarithm of the proportionate change in P is negatively related to lnP via $-\lambda$, note also the variance of the process is constant. Unlike Gibrat's model, economic forces are at work here, albeit in a simplistic fashion, since [8] may be re-written as a partial adjustment model with $(\eta+\mu t)/\lambda$ as the target or equilibrium city size and λ as the adjustment rate. In case [7] city size distributions are divergent through time and in case [8] city size distribution may be thought of as convergent or at least non-divergent in the sense that whilst both models predict increasing means [7] predicts increasing dispersion whereas [8] does not.

4.2 Tests for Distributional Form

Goodness of Fit (GF) Tests (Pearson, 1900) are a tool for examining the coherence of the data with a theoretical distribution. The tests are formed by dividing the range of a random variable into K mutually exclusive and exhaustive regions indexed k=1,.., K. Letting H_k be the hypothesized (or expected) number of observations falling in the k'th region (based upon sample size and the hypothesized distribution) and O_k be the number actually observed in that region, Pearson showed that f( ), the pdf of the test is asymptotically:

$$f(GF = \sum_{k=1}^{K} \frac{(H_k - O_k)^2}{H_k}) = \chi^2(k-1-j) \qquad [9]$$

where j corresponds to the number of parameters estimated in order to fully specify the null distribution. Small values of the test statistic correspond to coherence between observation and theory, large values to a lack of coherence, hence it is appropriate to employ a one sided upper tailed test.

Examining divergence is problematic. The standard "F" test for variance ratios is notoriously bad when the underlaying data generating processes diverge from normality (Anderson, 2001), furthermore divergence or polarization need not result in increased variances (Wolfson, 1994). The problem may be resolved by resorting to modifications of a test analogous to GF (herein denoted PAT for Pearson Analogue Test) used to examine the similarity between two samples (Anderson 2001, 2002). Letting O_{kA} be the observed proportion of observations in sample A falling in the k'th region, O_{kB} the corresponding value in sample B and J_k the proportion of observations falling in the k'th region when the samples are pooled, the test may be formulated as:

$$PAT = \frac{n_A n_B}{n_A + n_B} . \sum_{k=1}^{K} \frac{(O_{kA} - O_{kB})^2}{J_k} \qquad [10]$$

and it is distributed as $\chi^2(K\text{-}1)$. In this case small values of the test statistic correspond to coherence between the distributions underlaying the two samples and large values correspond to differences in the two underlaying distributions. When A and B correspond to time indices modifications of this test (essentially by employing particular linear combinations of elements in the sum, see Anderson (2002) for details) may be used to examine whether a distribution is diverging or polarizing over time whether due to the tails of the distribution moving apart or to there being increased concentration in them. To establish divergence when progressing from A to B the test has to be used twice, once to "not reject" divergence in moving from A to B and once to reject divergence from B to A. In all goodness of fit and divergence tests reported here partitions were chosen to generate 10 equiprobable cells or regions.

It has been common in empirical work in this field to interpret results of rank size regressions reported in Table 3 as evidence favoring the Pareto distribution and in particular as support for Zipf's law. However they do not constitute tests of the assumption that the data are generated by a Pareto distribution. Furthermore they only constitute an indirect test of Zipf's law in that, whilst the regression coefficient may not be significantly different from one, the underlying distribution may not be Pareto. As a consequence, following equation [3], more efficient estimates of the Pareto coefficients together with Goodness of Fit tests are reported in Panel A and B of Table 5 for China and the USA respectively. Since minimum city size has been set at 100,000 there is no need to estimate the lower bound, however maximum likelihood estimates are reported for information purposes.

[Table 5 about here]

With respect to the results for China the first thing to note is that the maximum likelihood estimates and the restricted regression estimates are consistently lower than the standard OLS estimates in Table 2 and uniformly lower than 1 in absolute value[10]. The Zipf's restriction ($\theta = 1$) is rejected in every instance except 1949. In addition the Goodness of Fit tests strongly reject the Pareto Distribution hypothesis in all but the 1949 observation set.

This is in striking contrast to the USA results presented in Panel B of Table 5 where although the Zipf restriction is rejected in 3 of 6 cases the Pareto distribution is never rejected at the 1% level and the coefficient estimates are uniformly higher than 1 in absolute value. Note also the estimated lower bounds are invariably higher than the true lower bound for China and invariably lower for the USA.

[Table 6 about here]

Double Pareto estimates for China are presented in Table 6. In the first three observation

[10] Note that when the Pareto parameter is less than one the variable has no finite moments.

years the information matrix is singular with the estimates converging to a Single rather than Double Pareto specification. (USA estimates always converged to the Single Pareto formulation reported in Panel B of Table 5). Apart from estimates of p* (which are closely related to the sample means), parameter estimates change little from period to period relative to their standard deviation in the remaining years. Again the Goodness of Fit tests strongly reject the Double Pareto hypothesis leaving the preponderance of evidence against the Double Pareto formulation for both China and the USA.

[Table 7 about here]

Panel A and B of Table 7 present the results for the log normal city size distribution specification for China and the USA respectively. With regard to China the Goodness of Fit Tests, with one exception (1980) fail to reject the null hypothesis of log normality at the 1% level, suggesting that the log-normal model is a much better rationalization of the data. The mean of the distribution indicates annualized percentage city size growth rates between successive observation periods of 0.73, -0.78, 1.84, 4.08, 4.34, 3.57, and 0.30 assuming a Gibrat process and 1.24, -0.45, 1.64, 2.87, 3.02, 2.66 and 0.47 assuming a Kalecki process. One aspect is perplexing for advocates of the Gibrat model, namely the diminishing estimates of the Standard Deviation over time from 1970 onwards. Under the Geometric Brownian Motion assumption the sample standard deviation is an estimate of $\sigma\sqrt{T}$ and should be increasing with the passing of time. As with the Pareto model, the results for the USA are in complete contrast to those for China with log-normality strongly rejected in every case though the standard deviation declines through time here as well.

[Diagrams about here]

The Chinese results are illustrated in diagrams for each observation year which overlay Kernel estimates of the empirical probability distribution function on respective plots of the estimated Pareto, Double Pareto and Log Normal distribution functions. Perusal of the diagrams makes it quite clear how it is that the log normal formulation best fits the data. Kernel[11] estimates of the distributions always generate modal points above the minimum city size (unlike the Pareto distribution) and the modal points always appear to be points of continuity (unlike the Double Pareto distribution). The close proximity of the Log Normal Distribution to the Kernel Density Estimate (relative to the other distributions) in each case attests to it providing the best fit. The issue remains as to whether convergence or divergence underlays the log normal city size data generating process for China, this is assessed in Table 8.

[Table 8 about here]

Assuming divergence in distribution to be reflected in increased variances a simple "F" test is appropriate, column 2 of Table 8 reports this. As may be observed the test is unable to detect any discernable change in variance from period to period except for 1985 to 1990 and 1990 to 1994 when variance reductions were indicated at the 5% level. As noted earlier this test has notoriously bad size and power properties and divergence need not imply variance change. Columns 3 through 6 of Table 8 report the details of divergence tests of the first and second order which, when considered at the 5% critical level, indicate no divergence or convergence from 1949 through to 1980 and convergence of both types (tails getting closer together and appearing less concentrated) from 1980 through to 1994 with divergence occurring from 1994 to 1999. It may be concluded that convergence appears to be the predominant force in the economic reform period suggesting that the Kalecki formulation rather than Gibrats formulation underlays the city size data generating process. Thus Chinese cities may be thought of as converging to an equilibrium city size which is itself trending upward through time.

---

[11]Kernel estimates were based upon the Epanechnikov kernel (Silverman [25]).

5. Conclusions

Whilst data on Chinese city sizes yield results for commonly employed rank size regressions that accord with the conventional wisdom of a Pareto distribution, closer examination reveals the Chinese case to be substantially different. Whereas a parallel study of the USA data reiterates support for a power law in the form of a Pareto distribution, the same specification is strongly rejected in the case of China in favor of a Log Normal distribution. The two distributions are linked. Describing the progress of city sizes through time by a Geometric Brownian Motion engenders a Log Normal city size distribution at a given point in time. Gabaix (1999) demonstrates that when the Geometric Brownian Motion is constrained by a lower reflective boundary a Pareto city size distribution at a given point in time results, moreover the Pareto coefficient will equal 1 when the growth rate of new cities is smaller than the growth rate of city sizes.

It would be simple to conclude that the lower reflective boundary is ineffective in the Chinese case but the evidence casts doubt upon this. The log normal distribution that derives from a Geometric Brownian Motion process has an upward trending mean and variance (Gibrat's Law) following from the independence of the process from initial city size. The evidence here suggests a constant or declining variance in all but two short intervals (1961-1970 and 1994-1999) with the declines being more apparent in the reform period. This is more in line with Kalecki's specification of the city size process which admits dependence between growth rates and city sizes through time but still generates log normality in the cross sectional distribution however with a non-increasing variance. In this formulation there is an equilibrium city size that could be trending through time (the result of technological innovation etc).

A striking feature of the reform period is the emergence of new cities in the middle of the distribution, the consequence of incredibly rapid growth in the early life of new entrant cities. This is not a feature of the pre reform period where a steady, but statistically insignificant, increase in the variance of the distribution is observed and new entrants appear at the lower end of the size distribution. This contrast is more striking given the attenuated gap between observation periods in the reform period. Thus a structural change does appear to have occurred at the onset of the reform period.

The growth in the numbers of cities should also be acknowledged since a precondition for Zipf's law in Gabaix (1999) is that the growth in the number of new cities should not exceed the growth in city sizes. In the case of China it does throughout the observation period. All in all Chinese city size distributions appear to differ substantially from the conventional wisdom that is embodied in Zipf's Law and appear to have been affected by the period of economic reform.

References

Alperovich, G. and J. Deutsch, 1995, The appropriateness of the Pareto distribution using a generalized Box-Cox transformation function, Journal of Regional Science 35, 267-276.

Anderson, G. J., 1994, Simple tests of distributional form", Journal of Econometrics 62, 265-276.

Anderson G. J., 2001, The power and size of nonparametric tests for common distributional characteristics, Econometric Reviews 20, 1-30.

Anderson G. J., 2002, Toward an empirical analysis of polarization, Working paper, Economics Department, University of Toronto.

Auerbach, F., 1913, Das Gesetz der Belvolkerungskoncertration, Petermanns Geographische Mitteilungen 59, 74-76.

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

Brakman S. , H. Garretsen, C. Van Marrewijk and M. van den Berg, 1999, The return of Zipf: toward a further understanding of the rank size distribution, Journal of Regional Science 39, 183-213.

Chenery H.B. and M. Syrquin, 1975, Patterns of development 1950-1970 (Oxford University Press).

Dobkins L. and Y. Ioannides, 2000, Dynamic evolution of the U.S. city size distribution, in: J. M. Huriot and J. F. Thisse, eds., Economics of cities (Cambridge University Press) 217-260.

Eaton J. and Z. Eckstein, 1997, City and growth: theory and evidence from France and Japan, Regional Science and Urban Economics 17, 443-474.

Gabaix, X., 1999, Zipf's law for cities: an explanation, Quarterly Journal of Economics 114, 739-767.

Gibrat, R., 1930, Une Loi Des Repartitions Economiques: L'effet Proportionelle, Bulletin de Statistique General, France, 19, 469.

Gibrat, R., 1931, Les Înegalites Economiques (Libraire du Recueil Sirey, Paris).

Hsing, Y., 1990, A note on functional forms and the urban size distribution, Journal of Urban Economics 27, 73-79.

Ioannides, Y. M. and H. G. Overman, 2003, Zipf's law fro cities: an empirical examination, Regional Science and Urban Economics, 33, 127-137.

Kalecki M., 1945, On the Gibrat distribution, Econometrica 13, 161- 170.

Kolmogorov, A. , 1933, Sulla determinaazione empirica di una legge di distribuzione, Giornale dell'Instituto Italiano degla Attuari 4, 83-91.

Overman, H. G. and Y. M. Ioannides, 2001, Cross-sectional evolution of the U.S. city size distribution, Journal of Urban Economics, 49, 543-566.

Pareto, V., 1897, Cours d'Economie Politique (Rouge et Cie, Paris).

Pearson, K., 1900, On a criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can reasonably be supposed to have arisen from random sampling, Philosophical Magazine 50, 157-175.

Quah, D., 1993, Empirical cross-section dynamics in economic growth, European Economic Review 37, 426-434.

Reed, W. J., 2001, The Pareto, Zipf and other power laws, Economics Letters, 74, 15-19.

Rosen, K. T. and M. Resnick, 1980, The size distribution of cities: an examination of the Pareto law and primacy, Journal of Urban Economics 8, 165-186.

Sharma, S., 2003, Persistence and stability in city growth, Journal of Urban Economics, 53, 300-320

Soo, K. T., 2002, Zipf's law for cities: a cross country investigation, Mimeo, Centre for Economic Performance, London School of Economics.

State Statistical Bureau, 1990, Forty Years of Urban Development.

State Statistical Bureàu, 1991-2000, Chinese Urban Statistical Yearbooks.

Silverman, B. W., 1985, Density estimation for statistics and data analysis, monographs on statistics and applied probability, (Chapman and Hall).

Simon, H., 1955, On a class of skew distribution functions, Biometrica 42, 425-440.

Steindl, J., 1965, Random processes and the growth of firms (Hafner, New York).

Wolfson, M.C., 1994, When inequalities diverge, American Economic Review Papers and Proceedings 84, 353-358.

Zipf, G., 1949, Human behavior and the principle of last effort (Cambridge MA: Addison Wesley).

Table 1
Summary Statistics of City Size

Panel A: Cities in China

Year	1949	1961	1970	1980	1985	1990	1994	1999
Mean	47.9	56.4	56.4	64	67.7	73.9	78.7	82.3
Std deviation	60.7	75.9	77.2	81.5	76.5	73.4	71.5	87.9
Median	28.4	35.3	29.9	36.7	43.6	57.1	63.2	64.1
Minimum	10.1	10.2	10.2	10.2	10.1	10.2	11.97	10
Maximum	418.9	641.2	580.2	601.3	698.3	783.5	953	1127
Sample size	77	176	164	208	313	453	606	658
Number of cities	132	208	177	223	324	467	622	667

Panel B: Cities in U.S.

Year	1930	1940	1950	1960	1970	1980	1990	2000
Mean	39.2	47	47.3	43.6	40.5	33.5	32.6	30.3
S.d.	81.1	89.5	89.9	81.6	76.5	64.8	63.3	60.2
Median	16.6	19.3	18.7	19.4	17.7	16.9	17.2	17.3
Minimum	10	10.1	10.2	10	10	10	10	10
Maximum	693	745.5	789.2	778.2	789.6	707.2	732.3	800.8
Sample size	94	98	112	136	161	168	192	238

Table 2
Testing β convergence hypothesis

	Cross Section Nonlinear Least Square		
	1961-1999	1961-1980	1980-1999
Constant	0.05* (0.004)	0.046* (0.007)	0.075* (0.009)
β	0.009* (0.002)	0.009* (0.002)	0.013* (0.003)
Adjusted R-square	0.23	0.12	0.14
Observation	157	157	157

Note: standard error in parentheses. * significant at 5%.

Table 3
Rank Size Regression, China 1949-1999

Variable	Coefficient	Variable	Coefficient
lnA	8.40* (0.11)	lnp	-1.08* (0.03)
D_{1949}	-1.35* (0.20)	$D_{1949} * lnp$	0.021 (0.055)
D_{1961}	-0.29* (0.16)	$D_{1961} * lnp$	0.003 (0.04)
D_{1970}	-0.53* (0.16)	$D_{1970} * lnp$	0.033 (0.041)
D_{1985}	0.85* (0.15)	$D_{1985} * lnp$	-0.078* (0.037)
D_{1990}	1.88* (0.14)	$D_{1990} * lnp$	-0.194* (0.036)
D_{1994}	2.68* (0.14)	$D_{1994} * lnp$	-0.286* (0.035)
D_{1999}	2.43* (0.14)	$D_{1999} * lnp$	-0.216* (0.034)
Observations		2655	
Adjusted R-square		0.9	

Note: standard error in parentheses. * significant at 5%.

Table 4

Panel A: Average Decade Transition Matrix in Pre-reform Period (1949-1980)

Upper boundary	0.25	0.5	0.75	1	2	∞	Exit
0.25	0.3	0.35	0.19	0.04	0	0	0.12
0.5	0.03	0.47	0.27	0.06	0.1	0	0.07
0.75	0.02	0.03	0.46	0.26	0.18	0.02	0.03
1	0	0.05	0.06	0.29	0.46	0.07	0.07
2	0	0.04	0.03	0.03	0.64	0.14	0.12
∞	0	0	0.02	0	0.02	0.96	0
Entry	0.44	0.33	0.07	0.06	0.09	0.01	

Panel B: Average Decade Transition Matrix in Reform Period (1980-1999)

Upper boundary	0.25	0.5	0.75	1	2	∞	Exit
0.25	0.49	0.36	0.04	0	0.07	0	0.04
0.5	0.01	0.63	0.21	0.04	0.07	0.02	0.02
0.75	0	0.03	0.71	0.18	0.07	0	0.01
1	0	0.08	0.06	0.78	0.08	0	0
2	0	0.04	0.03	0.05	0.79	0.08	0.01
∞	0	0	0	0	0.06	0.94	0
Entry	0.06	0.15	0.25	0.2	0.33	0.01	

Table 5
The Rank Order (Single Pareto) Distribution Model

Panel A: the case of China

Year	Sample size	p_{minML}	θ_{ML}	Var(θ_{MLE})	θ_{OLS}	Var(θ_{OLS})	$\chi^2(8)$ GF
1949	77	10.008	0.865	0.0097	0.914	0	12.481
1961	176	10.172	0.776	0.0034	0.859	0	36.841
1970	164	10.188	0.803	0.0039	0.875	0	21
1980	208	10.141	0.707	0.0024	0.799	0	60.75
1985	313	10.038	0.638	0.0013	0.737	0	187.927
1990	453	10.168	0.583	0.001	0.679	0	446.89
1994	606	11.95	0.603	0.001	0.702	0	686.574
1999	658	9.985	0.537	0	0.626	0	822.213

Note: upper tail probabilities (1-F(χ^2)) are not reported in Table 3 since they are all substantially less than 0.01 except for the initial year 1949.

Panel B: the case of U.S.

Year	Sample size	p_{minML}	θ_{ML}	Var(θ_{MLE})	θ_{OLS}	Var(θ_{OLS})	$\chi^2(8)$ GF	1-F(χ^2)
1950	112	10.062	1.046	0.0098	1.017	0.0004	16.214	0.039
1960	136	9.961	1.073	0.0085	1.06	0.0004	9.294	0.318
1970	161	9.941	1.131	0.008	1.111	0.0003	9.248	0.322
1980	192	9.946	1.301	0.0101	1.264	0.0003	10.691	0.22
1990	168	9.97	1.321	0.0091	1.295	0.0002	11.75	0.163
2000	238	9.984	1.392	0.0081	1.368	0.0003	6.118	0.634

Note:

p_{minML}	Unbiased transformation of Maximum Likelihood Estimate of the Lower bound (see footnote 5).
θ_{ML}	Maximum likelihood estimate of Zipf parameter.
Var(θ_{ML})	Variance of maximum likelihood estimate.
θ_{OLS}	Restricted Least Squares Estimate of Parameter.
Var(θ_{OLS})	Variance of Restricted Least Squares Estimate.
χ^2(8)GF	Pearson Goodness of Fit Test (based upon 10 equiprobable cells).
1-F(χ^2)	Upper tail probability of Pearson Goodness of Fit Tes

Table 6
Double-Pareto Model, the case of China

Year	Sample Size	α_{ML}	β_{ML}	p^*_{ML}	p_{MEAN}	$\chi^2(6)$ GF
1949	77	0.875	456532.13	2.317	3.46	
1961	176	0.78	5451775.2	2.325	3.608	
1970	164	0.807	4410943.2	2.3272	3.567	
1980	208	1.468 (0.156)	1.7383 (0.1526)	3.673 (0.0492)	3.731 (0.0592)	785.558
1985	313	1.600 (0.130)	1.810 (0.132)	3.836 (0.037)	3.874 (0.044)	1109.588
1990	453	1.794 (0.259)	1.765 (0.122)	4.039 (0.052)	4.034 (0.033)	1590.113
1994	606	1.978 (0.251)	1.940 (0.113)	4.146 (0.043)	4.141 (0.026)	2117.663
1999	658	1.980 (0.217)	1.999 (0.114)	4.162 (0.034)	4.164 (0.026)	2266.62

Note: standard error in parentheses

Table 7
Goodness-of fit Test of Log-normal Model

Panel A: City Size Distribution in China

Year	Sample size	Mean	Std Dev	Std Error of mean	$\chi^2(7)$ GF	$1-F(\chi^2)$
1949	77	3.46	0.831	0.095	7.805	0.35
1961	176	3.608	0.844	0.064	8.886	0.261
1970	164	3.567	0.877	0.069	16	0.025
1980	208	3.731	0.853	0.059	20.846	0.004
1985	313	3.874	0.779	0.044	11.313	0.126
1990	453	4.034	0.702	0.033	17.6181	0.014
1994	606	4.141	0.648	0.026	5.023	0.657
1999	658	4.164	0.661	0.026	11.392	0.122

Panel B: City Size Distribution in U.S.

Year	Sample size	Mean	Std Dev	Std Error of mean	$\chi^2(7)$ GF
1950	112	12.4749	0.9126	0.0862	40.1429
1960	136	12.4411	0.8706	0.0747	51.6471
1970	161	12.3909	0.8454	0.0666	64.4037
1980	168	12.2756	0.7547	0.0582	69.2619
1990	192	12.2667	0.7356	0.0531	70.1875
2000	238	12.2298	0.7003	0.0454	90.8235

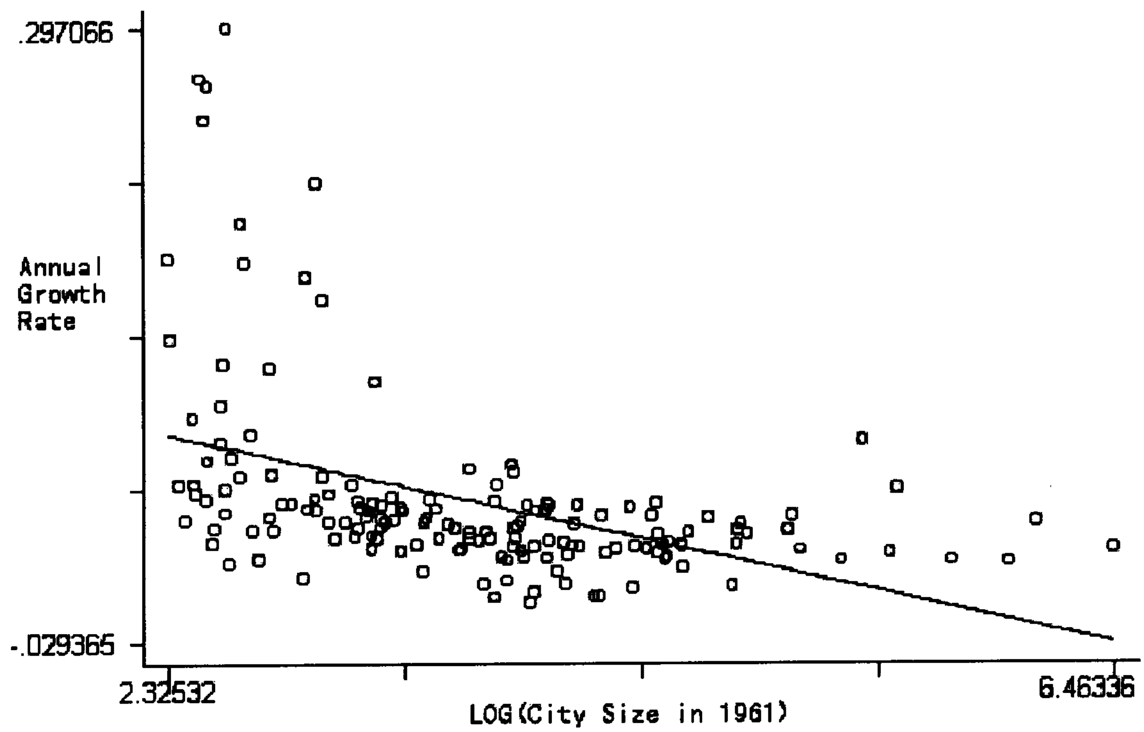
Note: upper tail probabilities ($1-F(\chi^2)$) are not reported in Table 5a since they are all substantially less than 0.01.

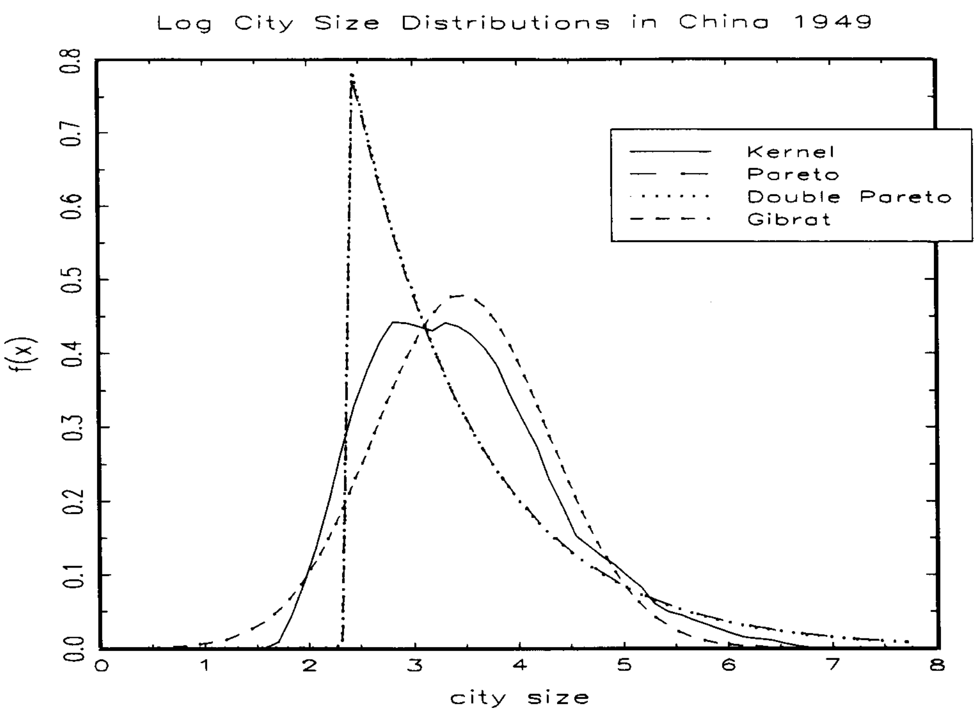
Table 8
Polarization Test

Comparison Years A/B	Variance Reduction Test Statistic and Upper Tail Probability[1]	A Diverging Relative to B[2]		B Diverging Relative to A[2]	
		Test Statistic	Upper Tail Probability	Test Statistic	Upper Tail Probability
49/61	0.9710	1.9749	0.8710	9.2695	0.1359
	0.5497	7.7241	0.2207	2.3386	0.7544
61/70	0.9248	12.106	0.0509	3.0532	0.7314
	0.6945	15.100	0.0199	21.255	0.0020
70/80	1.0168	0.1042	0.9995	2.8196	0.7653
	0.3524	1.3958	0.8448	2.4544	0.7312
80/85	1.1990	0.0000	1.0000	44.995	0.0000
	0.0740	0.0000	0.9555	44.995	0.0000
85/90	1.2313	0.0000	1.0000	54.121	0.0000
	0.0219	0.0000	0.9555	54.121	0.0000
90/94	1.1746	0.0000	1.0000	14.154	0.0245
	0.0329	0.0000	0.9519	14.154	0.0267
94/99	0.9619	6.5956	0.3110	0.0000	1.0000
	0.6864	6.5956	0.2890	0.0000	0.9008

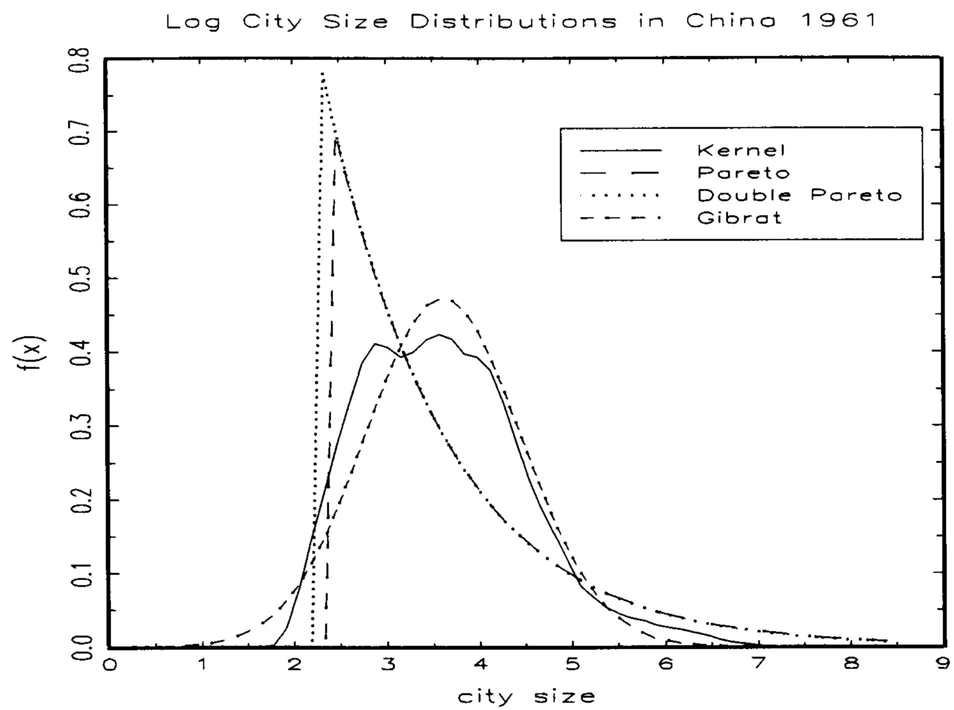
Notes:
1. Each cell of this column reports the standard variance ratio "F" test (with degrees of freedom (n_A-1, n_B-1) followed by its upper tail probability.
2. Columns 3 and 5 report the 1st and 2nd order polarization tests of Anderson (2002), columns 4 and 6 report the corresponding upper tail probabilities. 1st order polarization corresponds to tails moving further apart, 2nd Order Polarization refers to an increased concentration in the tails.

Figure 1. Annual Growth Rate and Initial City Size (1961-1999)

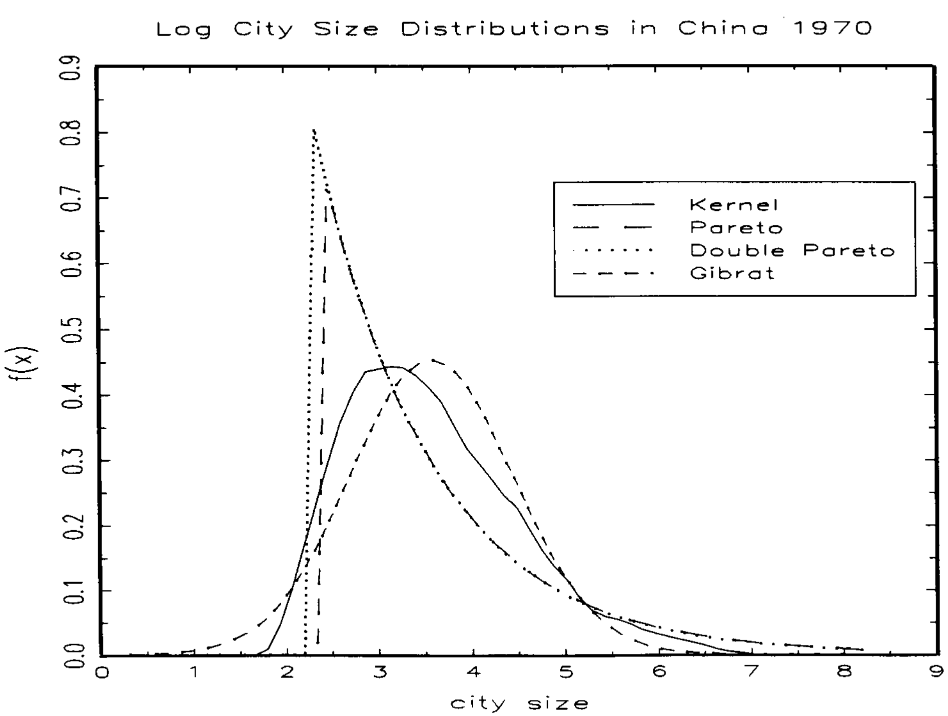




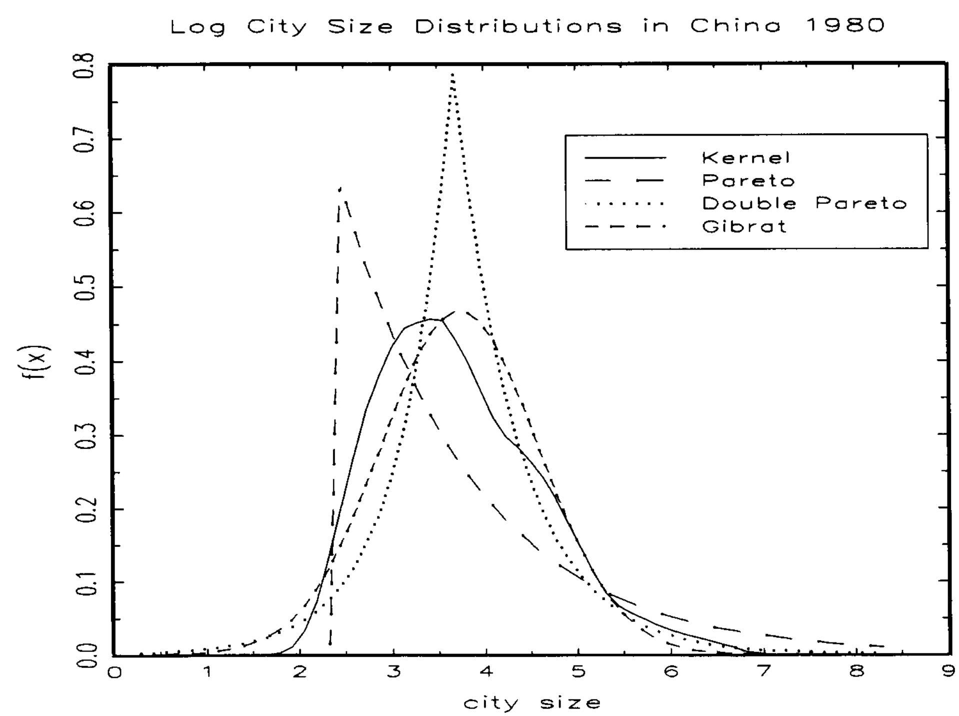
Log City Size Distributions in China 1949
f(x)
city size
Kernel
Pareto
Double Pareto
Gibrat



Log City Size Distributions in China 1961
f(x)
city size
Kernel
Pareto
Double Pareto
Gibrat

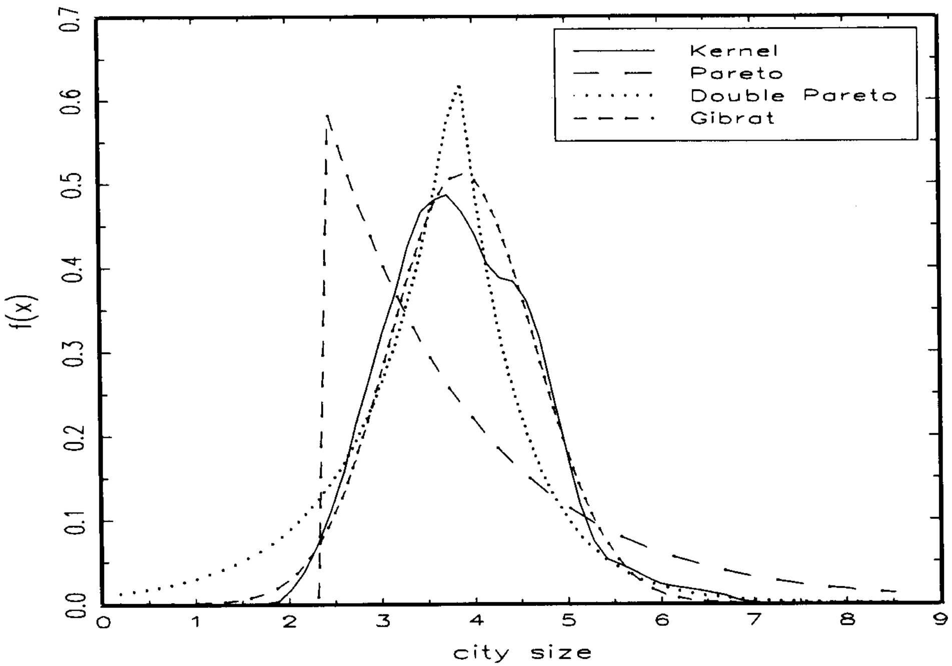


Log City Size Distributions in China 1970
f(x)
city size
Kernel
Pareto
Double Pareto
Gibrat

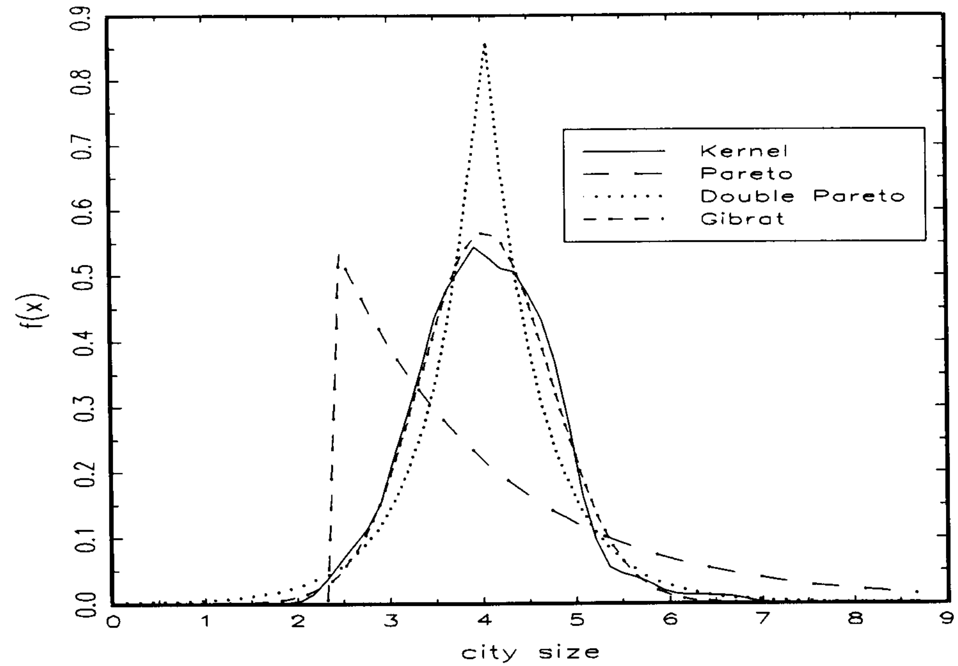


Log City Size Distributions in China 1980
f(x)
city size
Kernel
Pareto
Double Pareto
Gibrat

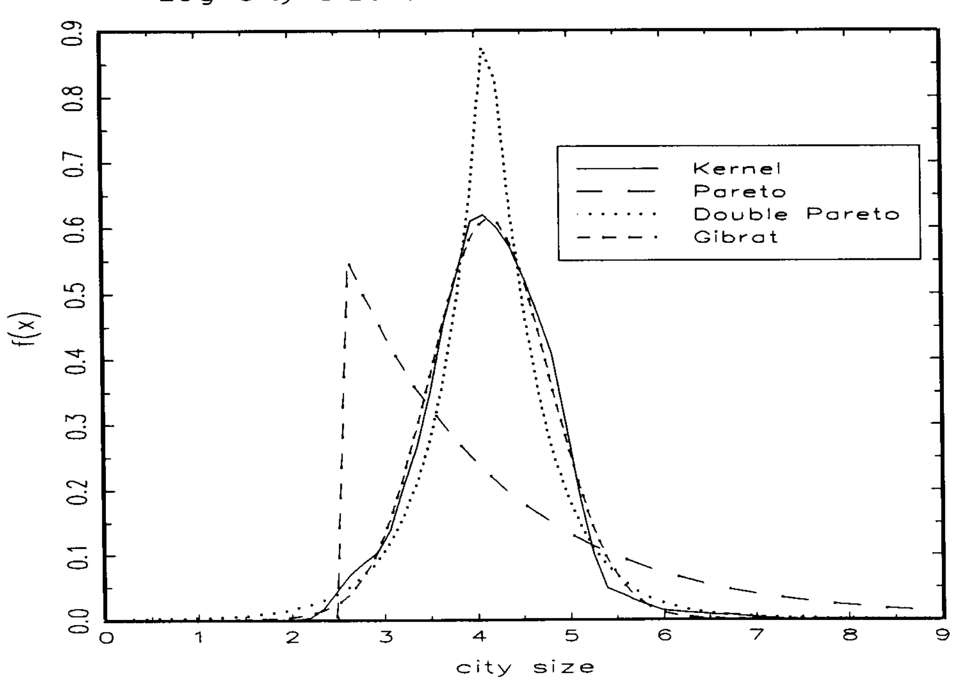
Log City Size Distributions in China 1985



Log City Size Distributions in China 1990



Log City Size Distributions in China 1994



Log City Size Distributions in China 1999