



HAL
open science

Rank-size city dynamics in China and India: 1981-2004

Alexandra Schaffar, Michel Dimou

► **To cite this version:**

Alexandra Schaffar, Michel Dimou. Rank-size city dynamics in China and India: 1981-2004. *Regional Studies*, 2010, pp.1. 10.1080/00343404.2010.521146 . hal-00648532

HAL Id: hal-00648532

<https://hal.science/hal-00648532>

Submitted on 6 Dec 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Rank-size city dynamics in China and India: 1981-2004

Journal:	<i>Regional Studies</i>
Manuscript ID:	CRES-2008-0313.R2
Manuscript Type:	Main Section
JEL codes:	J11 - Demographic Trends and Forecasts < J1 - Demographic Economics < J - Labor and Demographic Economics, R00 - General < R0 - General < R - Urban, Rural, and Regional Economics, R12 - Size and Spatial Distributions of Regional Economic Activity < R1 - General Regional Economics < R - Urban, Rural, and Regional Economics, R10 - General < R1 - General Regional Economics < R - Urban, Rural, and Regional Economics
Keywords:	Urban growth, City size distribution, Gibrat's law, Zipf's law, China, India

SCHOLARONE™
Manuscripts

Cris: note that the authors have provided the French abstract

Rank-Size City Dynamics in China and India: 1981-2004

ALEXANDRA SCHAFFAR* and MICHEL DIMOU†

* *PIMENT, University of La Reunion and LEAD, University of Toulon, 116 Rue General Ailleret, Campus Sud, Le Tampon, 97430, France.*
Email: alexandra.schaffar@univ-reunion.fr

† *PIMENT, University of La Reunion and LEAD, University of Toulon, 116 Rue General Ailleret, Campus Sud, Le Tampon, 97430, France.*
Email: dimou@univ-reunion.fr

(Received June 2008 : in revised form August 2010)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

SCHAFFAR A. and DIMOU M. By using empirical evidence based on data from India and China between 1981 and 2004, this paper studies the dynamic patterns of urban hierarchies within the two most populated countries in the world. The paper focuses on three specific issues: the distributional form of the rank-size relation, the urban growth patterns which do not reject the Gibrat' law for cities and, finally, the shifts in urban hierarchies for both countries over these two decades.

City-size distribution Urban growth Gibrat's law Zipf's law India China.

SCHAFFAR A. et DIMOU M. *Les dynamiques rang-taille des villes en Chine et en Inde : 1981-2004*. En s'appuyant sur la comparaison des évolutions des systèmes urbains chinois et indien entre 1981 et 2004, cet article propose une analyse de la formation des hiérarchies urbaines dans les pays en développement. De façon plus spécifique, trois axes de recherche sont abordés : la forme de la distribution rang-taille des villes, les caractéristiques de la croissance urbaine qui valide la loi de Gibrat et enfin les changements intra-distributionnels des villes, et par conséquent des hiérarchies urbaines, dans les deux pays.

Distribution rang-taille Croissance Urbaine Loi de Gibrat Loi de Zipf Inde Chine

JEL classification: R12, R10, J11, R23.

INTRODUCTION

Over the last twenty years, an extensive literature has developed on city size dynamics. This literature focuses on two main issues: the first concerns whether or not city size distribution follows Zipf's law, i.e. is the logarithm of city size a linear function of the logarithm of the city rank (MORICONI-EBRARD, 1993, GUERIN-PACE, 1995, GLAESER and al, 1995, KRUGMAN, 1996, EATON and ECKSTEIN, 1997, FUJITA and al, 1999, DOBKINS and IOANNIDES, 2000, CUBERES, 2004, GABAIX and IOANNIDES, 2004, ANDERSON and GE, 2005, SOO, 2005, DIMOU and SCHAFFAR, 2007), while the second concerns testing whether Gibrat's law holds for city growth, i.e. does city growth depend on city size (BLACK and HENDERSON, 2003, GABAIX and IOANNIDES, 2004, BOSKER and al, 2008).

In this literature, most studies have focused on evolution in the urban hierarchies of developed countries. This paper aims to deliver some empirical contribution on urban growth dynamics for economically emerging countries, by comparing city size evolution in China and in India, between 1981 and 2004. During this period, both countries turned away from State-owned and State-ruled economies towards a market economy and opened to international competition and foreign direct investment. In India, these changes took place in a rather progressive manner, without any major institutional shock affecting the country's demographics. In contrast, in China, a series of economical reforms have considerably eased Hukou

1
2
3 system rules¹ and abolished the prevailing cross-region labour mobility restrictions,
4
5
6 except for the biggest metropolitan areas. This resulted in boosting interregional
7
8
9 migration flows.

10
11
12 In order to determine whether or not these changes affected urban hierarchies
13
14 and dynamics of these two countries, two issues are examined here: firstly, does city
15
16 size distribution in China and India follow Zipf's law and, if this is the case, how has
17
18 this distribution evolved over the last twenty years? Secondly, in these two countries,
19
20 this distribution evolved over the last twenty years? Secondly, in these two countries,
21
22 is city growth a random or a deterministic process, sensitive to city size effects? In the
23
24 latter case, we focus more on city size growth characteristics and on the intra-
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
distributonal mobility of cities.

33 The paper is organised as follows. Section 2 investigates theoretical aspects by
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
presenting, firstly, recent discussions about the distributonal form of the rank-size
relation for cities and, secondly, the main approaches regarding city growth. Section
3 deals with methodology issues and data specification for the two countries' urban
systems. Section 4 tests the distributonal forms of the Chinese and Indian rank-size
relations for cities. Section 5 provides information on the shifts in urban hierarchies
in China and in India during the last two decades. Section 6 investigates city size
growth characteristics using panel unit root tests and Markov chains. The last section
concludes.

THEORETICAL ISSUES

Most researchers on urban hierarchies consider that city size distribution follows a Pareto principle. This is a very convenient assumption as it allows the study of urban hierarchies using a single parameter - the Pareto exponent. However, in recent years, some studies have questioned such an assumption, especially when one takes into account the entire distribution of city sizes and not just the upper truncated part.

ROSEN and RESNICK (1980) were among the first authors to clearly focus on a possible deviation from strict linearity between the logarithms of the rank and the size of cities. By working on a sample of 44 countries, they found evidence that, in most cases, the relation between city-rank and city size best fits a parabolic-type curve that can be studied by adding a quadratic term to Zipf's equation. SOO (2005) tested ROSEN and RESNICK's equation on the city size distribution of 73 countries: 2/3 of the countries had a quadratic term significantly higher (convexity) or lower (concavity) than 0, while only 1/3 of the sample's countries had a quadratic term significantly close to 0 (for which the Pareto law holds)².

While ROSEN and RESNICK studied a possible deviation from Zipf's law, some authors consider that if the entire population of cities is taken into account, then city size distribution is not a Pareto but a log-normal distribution. Following SIMON's (1955) earlier work, PARR (1985) assumed that only the upper truncated distribution

1
2
3 of city sizes follows a Pareto principle, while the entire distribution better fits a log-
4
5 normal distribution. However, he pointed out significant data and definition
6
7 problems that may occur in the latter case. Similar work on US and French city size
8
9 distributions by MALECKI (1980) and GUERIN-PACE (1985) revealed contradictory
10
11 evidence on the form of city size distributions with regard to the city size threshold.
12
13 However, both authors rejected the assumption that higher city size thresholds tend
14
15 to confirm a Pareto distribution, while lower ones reject it.
16
17
18
19
20
21
22
23

24 HSING (1990) claimed that in most economic literature, researchers take it for
25
26 granted that city size distribution follows a Pareto principle. He suggested using the
27
28 Box-Cox transformation in order to test log-linearity between city size and city rank.
29
30 By working on different data samples, HSING (1990) often rejected the fit between
31
32 the US city size distribution and the Pareto law. Following HSING, CAMERON
33
34 (1990) and ALPEROVITCH and DEUTSCH (1995) came to the same conclusions,
35
36 when studying the city size distributions of twenty five different countries.
37
38
39
40
41
42
43

44 In a more recent work, EECKHOUT (2004) delivered a rigorous study of the
45
46 relation between city rank and city size for 23,539 US settlements in 2000
47
48 (representing 75% of the US population). Using a Kolmogorov-Smirnov test,
49
50 EECKHOUT proved that the distribution of city sizes better fits a log-normal
51
52 distribution, with a mean (in log) equal to 7.28 and a standard deviation (in log)
53
54 equal to 1.75. According to EECKHOUT, most studies building on the US city size
55
56
57
58
59
60

1
2
3 distribution accept the validity of Zipf's law because they work on truncated
4
5 distributions. KRUGMAN (1996), GABAIX (1999) or BLACK and HENDERSON
6
7 (2003), all used small samples of about 135 cities, which corresponds to 0.5% of US
8
9 settlements and 30.2% of US population. This truncated distribution fits the tail of the
10
11 log-normal distribution for the entire population of cities.
12
13
14
15
16
17

18 Following EECKHOUT (2004), some studies (ANDERSON and GE, 2005,
19
20 ZEYNELOGLU and al, 2005, NOTA and SONG, 2007) have confirmed the fit
21
22 between the log-normal and the entire city size distribution and reject an arbitrary
23
24 choice for any city size threshold. However, their definition of a city is problematic,
25
26 since there seems to be no boundary between an urban area and a rural one, right
27
28 down to a single isolated household. In EECKHOUT's (2004) sample, the median city
29
30 size is of 1337 inhabitants, which means that half of the settlements can be considered
31
32 as merely small villages. This contrasts with the fact that 80% of the US urban
33
34 population lives within the 256 largest urban agglomerations of the country. One
35
36 could therefore easily argue that a truncated distribution reveals urban hierarchy
37
38 patterns much better than the entire one. Moreover, the log-normal distribution
39
40 depends upon three parameters, that is, the mean city size, its standard deviation
41
42 and the number of cities, which leads to a less convenient interpretational framework
43
44 than the Pareto one, which depends upon a single parameter, the hierarchy
45
46 exponent.
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 The distributional form of the relation between the ranks and the sizes of cities
4 seem strictly related to city growth dynamics. Many studies investigate what should
5 be the nature of urban growth in order to lead to a steady distribution that follows
6 either Zipf's law or an alternative theoretical distribution such as the log-normal or
7 the double Pareto (REED, 2001).
8
9

10
11
12
13
14
15
16
17
18 City growth theories can roughly be divided into random growth (GABAIX,
19 1999, GABAIX and IOANNIDES, 2004, CORDOBA, 2008) and deterministic growth
20 theories (EATON and ECKSTEIN, 1997, BLACK and HENDERSON, 2003,
21 HENDERSON and WANG, 2007) although, in recent years, some attempts have been
22 made to provide a less dichotomous theoretical vision (ROSSI-HANSBERG and
23 WRIGHT, 2007, DURANTON, 2006 and 2007).
24
25
26
27
28
29
30
31
32
33

34
35
36 Random growth theories consider that city growth is a stochastic process which,
37 in the steady state, produces, for large cities, a rank-size distribution that obeys Zipf's
38 law. These theories follow GABAIX's (1999) basic model, developed under the very
39 restrictive hypotheses of a growing population, free labour mobility for young
40 households and constant returns to scale technologies. The growth of cities appears
41 to be a random walk process, linked to randomly distributed exogenous shocks,
42 which generate urban amenities in a multiplicative way and subsequently modify a
43 city's household utility function. These amenity shocks are either policy shocks
44 (related to the level of taxes, to pollution control, to the quality of public and
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 municipal services and infrastructure such as roads and schools, or to environment
4
5 protection measures) or natural and historical shocks (due to natural catastrophes,
6
7 wars, diseases or poor harvests in less developed economies), but GABAIX also
8
9 considers the possibility of industry-specific productivity shocks in bigger cities.
10
11
12

13
14
15 At the opposite end of the spectrum, deterministic growth theories assume that
16
17 cities' differing sizes depend on firms' location choices, made when comparing the
18
19 advantages and drawbacks of each city. Firms concentrate geographically in order to
20
21 take advantage of conglomeration effects, linked either to specialisation (MAR
22
23 externalities) or to diversification (Jacob externalities), but suffer, on the other hand,
24
25 from diseconomies due to congestion and the costs of commuting. Following this
26
27 theoretical framework, EATON and ECKSTEIN (1997) and BLACK and
28
29 HENDERSON (1999, 2003) developed endogenous growth models in which the size
30
31 of a city depends on parameters such as scale externalities, human capital
32
33 concentration and localised information spillovers. Deterministic growth theories do
34
35 not necessarily produce Zipf's law.
36
37
38
39
40
41
42
43
44

45
46
47 BLACK and HENDERSON (1999) tested the nature of urban growth, while
48
49 taking into account the heterogeneity of cities through differentiated stocks of human
50
51 capital. In their initial work on the evolution of the US city size distribution between
52
53 1900 and 1990, they found that cities exhibited parallel growth patterns. However, by
54
55 relaxing some hypotheses, in particular the strong relationship between urban
56
57
58
59
60

1
2
3 growth and the role of education to which the growth in human capital was almost
4
5 exclusively linked, BLACK and HENDERSON (2003) found, in a more recent work,
6
7
8 that the average US city size increased under the impetus of technological changes
9
10 and human capital accumulation, but also that the smaller cities grew faster than the
11
12 large ones, hence leading to a convergence towards an optimal city size.
13
14
15

16
17
18 Many studies have tried to test the nature of urban growth in order to explain the
19
20 steady distributional form of city rank versus city size. When studying growth in
21
22 France and Japan in the 20th Century, and working with small samples, EATON and
23
24 ECKSTEIN (1997) found that the k largest cities in each country maintained their
25
26 ranking over the entire reference period, which means that city growth is parallel,
27
28 rather than convergent or divergent. On the other hand, in order to test the random
29
30 growth hypothesis, IOANNIDES and OVERMAN (2003) produced a non-parametric
31
32 estimation of urban growth, by using a sample of large metropolitan areas in the
33
34 United States from 1900 to 1990. They found that the growth process of US cities
35
36 obeys Gibrat's law and hence scale effects play no determining role in city growth
37
38 dynamics.
39
40
41
42
43
44
45
46
47
48
49

50 Some researchers have focused on conflict-affected distributions, in order to test
51
52 the consequences of random isolated shocks to the relocation of labour and to the
53
54 stability of the city size distribution over time. DAVIS and WEINSTEIN (2002) and
55
56 BOSKER and al (2008) examined the time effects of Second World War Allied
57
58
59
60

1
2
3 bombing on German and Japanese city distributions, with contradictory results.
4
5
6 SHARMA (2003) rejected Gibrat's law when studying the long term effects of
7
8
9 partition on India's urban hierarchies, with respect not only to the city sizes but also
10
11 to the cities' relative growth. Empirical studies on urban growth dynamics in
12
13 Portugal, the US or in the Balkans also reveal departures from Gibrat's law
14
15 (DELGADO and GODINHO, 2007, GARMESTANI and al, 2007, DIMOU and
16
17 SCHAFFAR, 2009).
18
19
20
21
22
23

24 While empirical work delivers contradictory evidence when it comes to the
25
26 nature of urban growth and to the distributional patterns of the rank-size relation,
27
28 some recent theoretical contributions from ROSSI-HANSBERG and WRIGHT (2007)
29
30 and DURANTON (2007) highlight a less deterministic approach. Following
31
32 IOANNIDES and OVERMAN's work, ROSSI-HANSBERG and WRIGHT (2007)
33
34 introduce a model able to predict random urban growth under some restrictive
35
36 conditions such as the elimination of the effects of human capital. ROSSI-
37
38 HANSBERG and WRIGHT (2007) admit that city growth depends upon randomly
39
40 distributed productivity shocks, specific to a localised industrial sector or group of
41
42 firms. This however does not necessarily result in Zipf's law, as predicted by random
43
44 growth theories, because of the short-term effects of the exogenous productivity
45
46 shocks. It also supposes that cities are specialised in a single industry, which is rarely
47
48 the case.
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4 In a similar way, DURANTON (2006 and 2007) modelled the micro-foundations
5
6 of the stochastic process which leads, through industrial churning, product
7
8 proliferation and local spillovers, to urban growth. Under the assumption of
9
10 monopolistic competition, free movement of labour and a fixed number of cities,
11
12 DURANTON showed that innovation-driven shocks at the level of industries and
13
14 cities may explain the growth and decline of cities. When introducing agglomeration
15
16 economies and congestion, the model predicts some convergence, leading to a right
17
18 skewed (concave) rank-size distribution of cities, which reproduces various French
19
20 and US features, observed over the last twenty years.
21
22
23
24
25
26
27
28

29
30 By studying the dynamics of Chinese and Indian urban hierarchies from 1981 to
31
32 2004, this paper aims to test the validity of these theories within two countries that
33
34 have had high economic growth. If city growth is independent from city size, that is
35
36 if Gibrat's law for cities holds, then urban change is not affected by scale effects.
37
38 Conversely, if city size growth is a deterministic process, one should examine
39
40 whether this leads to convergence towards an optimal city size, with small cities
41
42 growing faster than the bigger ones; divergence, with demographic concentration
43
44 towards metropolitan areas; or parallel growth, where cities keep their initial
45
46 hierarchical ranks during the whole period.
47
48
49
50
51
52
53
54
55
56
57
58
59
60

METHODOLOGY AND DATA ISSUES

From a methodological point of view, this paper first examines the distributional patterns of the rank-size relation for Chinese and Indian cities, then studies their changes on a decennial base, and finally focuses on urban growth processes and city growth dynamics.

To begin with, we use a Kolmogorov-Smirnov test to examine the distributional patterns of the rank-size relation, with respect to two theoretical distributions, the Pareto and the log-normal. The null hypothesis H_0 is that the empirical distribution F follows a theoretical law F_0 , while the alternative hypothesis H_1 is that the two distributions do not match. The goodness-of-fit test provides evidence that the Chinese and Indian city size distributions, when truncated, are Pareto.

Secondly, we estimate the Pareto coefficient for the 1984, 1994 and 2004 Chinese city size distributions and for the 1981, 1991 and 2001 Indian distributions. In both cases, we use the GABAIX and IBRAGIMOV (2009) correction which considerably reduces the bias for the OLS method when applied to small finite samples:

$$\ln(R_i - \xi) = \ln A - \beta \ln S_i$$

with $0 \leq \xi < 1$. When $\xi = 0$, Zipf's law holds perfectly. According to GABAIX and IBRAGIMOV (2009), the best estimation of β , for small samples, is provided when $\xi = 1/2$.

$$\ln(R_i - \frac{1}{2}) = \ln C - \hat{\beta} \ln S_i$$

with the approximate true standard error of the estimated coefficient given by $\xi \sqrt{2/n}$ (KRATZ and RESNICK, 1996, GABAIX and IBRAGIMOV, 2006, DIMOU and SCHAFFAR, 2009).

Thirdly, we investigate urban growth patterns. We first test for stationarity in city size. We considered that a city's sizes are generally correlated over time, because of the durability of the publicly-owned structures, houses and real estate investments. If $\ln(S_{it})$ is the logarithm of the population of a city i at time t , then the size of a city is a first order auto-correlation process such as:

$$\ln S_{i,t} = \phi_i \ln S_{i,t-1} + \varepsilon_{it}$$

where $\phi_{i,t}$ is the first order regression coefficient and ε_{it} an isolated shock at time t (DIMOU and SCHAFFAR, 2009). We use two specifications of the model, both with fixed effects and with or without drift:

$$\Delta \ln S_{i,t} = \alpha + \gamma_i \ln S_{i,t-1} + \sum_{j=1}^k \rho_{ij} \Delta \ln S_{i,t-j} + \varepsilon_{it}$$

$$\Delta \ln S_t = \alpha + \theta_i t + \gamma_i \ln S_{i,t-1} + \sum_{j=1}^k \rho_{ij} \Delta \ln S_{t-j} + \varepsilon_{it}$$

where $\gamma_i = \phi_i - 1$, α is a constant, $\theta_i t$ control for the upward trending and k is the number of the lagged difference term for city i . $H_0 : \gamma_i = 0$ ($\phi_i = 1$) is the instability hypothesis, where cities' sizes are non-stationary, versus the alternative $H_1 : \gamma_i < 0$ ($\phi_i < 1$), where the logarithms of cities' sizes converge to a constant value in the steady state. The null hypothesis does not reject Gibrat's law for cities.

As SHARMA (2003) and HURLIN and MIGNON (2005) have pointed out, the commonly-used unit root tests, such as the Dickey-Fuller (DF) and the Augmented Dickey-Fuller (ADF) show little power in distinguishing the unit root from meaningful stationary alternatives, when applied to a single equation for short time series. Panel unit root tests appear to be more efficient. The main limitation of these tests is that they are all constructed under the assumption that the individual time series in the panel are cross-sectionally independently distributed (PESARAN, 2007). When it comes to city size dynamics, such an assumption may be spurious: for example, urban growth may be affected by some common macroeconomic factors for cities belonging to the same region or State; or interregional migration flows may commonly affect some cities closely located. Second generation panel unit root tests abandon the restrictive assumption of the previous tests and model cross-section dependence.

1
2
3
4 Previous empirical literature on urban growth only delivered first generation
5
6 panel unit root tests. This paper performs both the LEVIN and al (2002) and IM and
7
8 al (2003) first generation tests and the CHOI (2002) and PESARAN (2007) second
9
10 generation tests. While the first ones do not come to any clear conclusions, the second
11
12 ones provide evidence that city sizes are not stationary during the 1981–2004 period,
13
14 for both countries.
15
16
17
18
19
20

21 Finally, we study intra-distributional dynamics for Chinese and Indian cities.
22
23 Following EATON and ECKSTEIN (1997), DOBKINS and IOANNIDES (2000),
24
25 BLACK and HENDERSON (2003), we use Markov chain techniques. This requires
26
27 the discretisation of the distribution, by assigning each city to one of a predetermined
28
29 number of groups, based on its relative size (BOSKER and al, 2008). In this paper, the
30
31 discretisation process draws upon QUAH (1993) and EATON and ECKSTEIN (1997),
32
33 with cut-off points exogenously defined at relative city sizes of $0.3m$, $0.5m$, m , $2m$,
34
35 with m the average city size, for a given year t . We assume that the distribution
36
37 follows a homogenous stationary first order Markov process and F_t denotes the
38
39 vector of distributional shares for each group of the discretised distribution:
40
41
42
43
44
45
46
47
48
49

$$F_{t+1} = M \times F_t$$

50
51
52
53 Each element p_{ij}^t of the transition matrix M^t represents the probability that a city
54
55 moves from the group i to the group j in t . The transition probabilities are estimated
56
57
58
59
60

1
2
3 with the maximum likelihood method. If $p_{ij}^t \neq 0$, for all i, j then the Markov chain is
4
5
6 ergodic, which means that any city from any group may attain any other group of
7
8
9 the distribution. The Markov chain techniques deliver evidence for some
10
11
12 differentiation between Chinese and Indian urban dynamics. In India, city size intra-
13
14
15 distributional dynamics feature stability, while in China more significant changes
16
17
18 occur.

19
20
21 With regard to data issues, in this work, cities are defined from an administrative
22
23
24 point of view. This allows us to obtain long-run annual city size series³. The city size
25
26
27 threshold of our samples is 100,000 inhabitants, which is also the threshold used by
28
29
30 official Chinese and Indian statistical services for city data, although in China, this
31
32
33 definition has changed over time. Cities above this threshold represent 59.01% of the
34
35
36 urban population in China and 69.25% in India⁴. One can therefore easily consider
37
38
39 that both samples represent quite well the respective national urban systems.
40
41
42 Furthermore, annual data for cities under 100,000 inhabitants is not always available.

43
44
45 Chinese data is provided by the Chinese Urban Statistical Yearbooks. The city
46
47
48 size series go from 1984 to 2004 and obey the 1984 criteria for defining a city,
49
50
51 adjusted to year 2000 statistical changes (ZHOU and MA, 2003, XU LAI, 2008). Hong
52
53
54 Kong and the Tibetan cities are excluded. Table 1 gives information about the
55
56
57 Chinese sample.

58
59
60 Table 1

1
2
3
4 Three comments can be made: firstly, the number of cities above 100,000
5
6 inhabitants has grown strongly in China during this 20-years period (the annual
7
8 growth rate of cities entering the sample is 4.82%). However, the trend was faster
9
10 during the 1984-1994 decade (with an annual rate of entry of 7.69%) than the 1994-
11
12 2004 period (with an annual rate of 2.03%). Secondly, when comparing mean, median
13
14 and quartile values for city size, except for the first quartile, all other values decrease
15
16 between 1984 and 1994 and increase afterwards, which indicates a significant change
17
18 in urban growth trends: during the first decade, urban growth mainly concerns cities
19
20 under 250,000 inhabitants, while, during the second decade, growth is higher for
21
22 cities above 370,000 inhabitants. Thirdly, the medial city size strongly falls during the
23
24 first decade, then increases, but does not get back to its 1984 level. This confirms the
25
26 previous comments and also illustrates the results of repeated anti-megacity policies
27
28 implemented by successive Chinese governments.
29
30
31
32
33
34
35
36
37
38
39

40 Table 2 provides the same type of information for the Indian sample in 1981, 1991
41
42 and 2001, which are the Census years. Cities from Jammu and Kashmir and Annam
43
44 are not included, because the time-series were incomplete.
45
46
47
48
49

50 Table 2
51
52

53 The number of cities with a size over 100,000 inhabitants almost doubled between
54
55 1981 and 2001 but, contrary to the Chinese case, growth trends seem to have been
56
57 particularly steady: the annual growth rate of cities entering the sample was 3.25%
58
59
60

1
2
3 and remained the same during the whole period. This entry rate is very close to the
4
5
6 annual growth of the urban population (3.01%), which could be explained by the fact
7
8
9 that urban demographics in India mainly depend upon the natural increase in
10
11
12 population (DATTA, 2006).
13

14
15 In parallel, the mean city size slowly increased (from 361,500 to 403,000
16
17 inhabitants) and remained determinedly higher than during the third quartile, which
18
19 means that India's urban system is characterised by the presence of some over-
20
21 populated metropolises and a large group of small and medium-size cities. The study
22
23 of the medial city size and its evolution confirms the above trends: in 2001, 41 cities
24
25 were home to 50% of the urban population, compared with 24 cities in 1981, while
26
27 the difference between the median and the medial city size increases. This could be
28
29 interpreted as a sign of non-convergence in city size over time.
30
31
32
33
34
35
36
37
38
39

40 CHINA'S AND INDIA'S RANK-SIZE DISTRIBUTIONAL 41 42 43 PATTERNS 44 45 46 47 48 49

50 Most studies consider that the upper truncated part of the Chinese city size
51
52 distribution follows a Pareto law (SONG and ZHANG, 2002). In a recent work, GAN,
53
54 LI and SONG (2006) contested this, while ANDERSON and GE (2005) used the
55
56 Pearson test to argue that the log-normal distribution provides a better fit for Chinese
57
58 cities with a size above 100,000 inhabitants. However, their definition of a city is
59
60

1
2
3 based on the 1963 criterion and strongly differs from the actual definition, so the
4
5 samples also differ. No corresponding study exists for Indian city-size distribution.
6
7
8
9

10 We provide a KS-test for the 1984, 1994 and 2004 Chinese city size distributions
11
12 and the 1981, 1991 and 2001 Indian city-size distributions. Histograms of city sizes in
13
14 Figure 1 seem to indicate that the Pareto law provides a better fit than the log-normal
15
16 distribution.
17
18
19
20
21

22 Figure 1
23
24
25

26 We first run the KS-test for a Pareto distribution. In order to optimize its
27
28 performance we use, for the Pareto exponent, values that range between 0.8 and 1.5.
29
30 Tables 3 and 4 provide the results for both countries.
31
32
33
34

35 Table 3
36
37
38

39 Figure 2
40
41
42

43 For China, at the 5% level, all KS-stats have lower values than the critical values
44
45 (p-values are high), which means that H_0 cannot be rejected: the distribution of
46
47 Chinese cities with sizes above 100,000 inhabitants is Pareto.
48
49
50
51

52 Table 4
53
54
55

56 Figure 3
57
58
59
60

1
2
3 India's city size distribution reveals an even better fit with the Pareto law. Results
4
5
6 in Table 4 show that the null hypothesis is never rejected, with high p-values, which
7
8
9 also means that the rank-size distribution of Indian cities of more than 100,000
10
11
12 inhabitants follows a Pareto principle.

13
14
15 We proceed by running the same tests for the fit with the log-normal distribution.
16
17
18 All KS-stats reject the null hypothesis, at the 1% level, when we use the estimated
19
20
21 parameters from our samples (Tables 5 and 6). It seems, however, more appropriate
22
23
24 to use the hypothesis that the distribution for cities with sizes above the threshold of
25
26
27 100,000 inhabitants is the upper part from a truncated log-normal distribution
28
29
30 (Figure 4) that takes into account all the cities, as EECKHOUT (2004) suggests.

31 32 33 Figure 4

34
35
36
37 The main problem when considering this issue is that the parameters of the log-
38
39
40 normal distribution cannot be estimated, so one needs to proceed by simulation. We
41
42
43 have run a large number of simulations to test this fit. Among them, we use, for the
44
45
46 Chinese case, the parameters estimated by ANDERSON and GE (2005) who find that
47
48
49 China's city-size distribution better fits a log-normal distribution and also calculate
50
51
52 new parameters, for example, by dividing all the non-agricultural population by the
53
54
55 number of urban and rural districts.
56
57
58
59
60

1
2
3 Table 5 provides the results for both tests in 2004. In all cases, the p-values remain
4 close to 0, which leads us to always reject the null hypothesis. The results are the
5 same for 1984 and 1994.
6
7
8
9

10
11
12
13 Table 5
14

15
16
17 We now test whether the Indian city-size distribution fits the log-normal one. The
18 null hypothesis is rejected when we consider the estimated parameters from the
19 sample, for all years. Then we also run different simulations. We use parameters
20 calculated by dividing the total urban population by the number of districts of above
21 5,000 inhabitants. The null hypothesis is rejected for 1981, 1991 and 2001 (all p-values
22 are close to 0) at the 1% level. Table 6 provides our findings for 2001. The 1981 and
23 1991 results are similar.
24
25
26
27
28
29
30
31
32
33
34
35
36
37

38 Table 6
39
40

41
42 Obviously, after running a great number of simulations (for example, by keeping
43 the urban population fixed and by varying the number of human settlements) we can
44 calibrate a log-normal distribution (with a very large number of settlements and a
45 very large standard deviation from mean size) that fits the empirical distributions of
46 Chinese and Indian cities. Such a fit seems however a fictive construction for both
47 countries' urban systems that cannot be verified.
48
49
50
51
52
53
54
55
56
57
58
59
60

SHIFTS IN CHINESE AND INDIAN URBAN HIERARCHIES

Following on from the conclusions in the previous section, we can now study the evolution of the Pareto exponents for the Chinese and Indian city-size distributions over time. Figures 5 and 6 draw the density kernels for each country and for different dates.

Figure 5

In Figure 5, the 1984 and 2004 density kernels for Chinese cities are quite similar, and unlike the 1994 one which is steeper. This reveals some convergence towards a unique city size during the first decade, but not during the second one. The Indian city-size distribution seems fairly steady over time. In Figure 6, the density kernels for 1981, 1991 and 2001 overlap. The long thin right tail of the kernel indicates the existence of some very large cities, which contrasts with massive presence of cities with less than 200,000 inhabitants.

Figure 6

We use the GABAIX-IBRAGIMOV method to estimate the Pareto exponents for both distributions. Table 7 provides the results.

Table 7

1
2
3
4 Three comments can be made. Firstly, the Pareto exponent is close to 1 for the
5
6 Chinese distribution but significantly higher than 1 for the Indian distribution. The
7
8 latter appears less hierarchical than the former. Secondly, the trend of the Pareto
9
10 exponent confirms the contrasted movement of the Chinese distribution during the
11
12 last two decades (its value is 1.076 in 1984, 1.198 in 1994 and 1.089 in 2004) and the
13
14 slow but progressive movement towards concentration in Indian cities (its value
15
16 steadily decreases from 1981 to 2001). Thirdly, Zipf's law (the H_0 hypothesis) holds
17
18 better for the Chinese distributions than for the Indian ones. The results given in
19
20 Table 7 are close to SONG and ZHANG's (2002) findings, although they use the
21
22 simple OLS method for estimating the Pareto exponent, which leads them to always
23
24 reject Zipf' law. ANDERSON and GE's (2005) findings also follow the same trend but
25
26 are slightly higher, because of their definition of cities.
27
28
29
30
31
32
33
34
35
36

37
38 Following GUÉRIN-PACE (1995) and SONG and ZHANG (2002), we now
39
40 examine shifts in both countries' urban hierarchies when the city-size threshold
41
42 changes. Table 8 provides results for Chinese city distributions, with city-sizes above
43
44 0.1, 0.2, 0.5 and 1 million inhabitants.
45
46
47
48
49

50 Table 8
51

52
53 The Pareto exponent increases along with the city-size threshold. This implies
54
55 that urban hierarchy is less pronounced when one considers only the upper part of
56
57 the rank-size distribution of the Chinese cities. Zipf's law holds only for the largest
58
59
60

1
2
3 samples, i.e. the ones that include all cities with a population above 100,000
4
5
6 inhabitants.

7
8
9
10 When it comes to Indian cities, the trends are less clear (Table 9). Whatever city-
11
12 size threshold we choose, Zipf's law is rejected, with the Pareto exponent values
13
14 being strictly higher than 1. Close inspection of these results yields evidence of some
15
16 specific shifts in Indian urban hierarchies. From 1981 to 1991, the Pareto exponent
17
18 decreases for the whole sample, which means that urban concentration increased.
19
20 This completely changes when we take into account only cities with more than 1
21
22 million inhabitants. India's city size distribution tends to be more hierarchical, except
23
24 for the group of the largest cities.
25
26
27
28
29
30
31
32

33 Table 9

34
35
36
37 The results above allow us to draw two comparative conclusions concerning the
38
39 Chinese and Indian city-size distribution shifts over the last two decades.
40
41
42

43
44 Firstly, the two countries' urban hierarchies differ: India's city size distribution is
45
46 characterised by the presence of a large number of small cities and some very large
47
48 metropolises such as Bombay, Delhi and Calcutta. Zipf's law is rejected for any city
49
50 size threshold. When it comes to the Chinese distribution, Zipf's law holds only for
51
52 the whole sample. The distribution gets less hierarchical than what is predicted by
53
54 the Zipf's law, when city size threshold increases.
55
56
57
58
59
60

1
2
3
4 Secondly, the two countries' city-size distributions follow different trends. In
5
6 China, the number of small cities strongly grows until the middle of the 1990s,
7
8 followed, during the last decade, by an increase in the number of medium-size cities.
9
10 The larger cities' distribution remains under the effects of anti-megacities policies.
11
12 On the other hand, India's city-size distribution tends to become more hierarchical.
13
14
15
16
17
18
19

20 **CHINESE AND INDIAN URBAN GROWTH PATTERNS**

21
22
23
24

25 In order to study urban growth characteristics in China and India, we run panel
26
27 unit root tests for non stationarity in city size. In the presence of a unit root, the
28
29 underlying trend is stochastic, which implies that a series has a long memory and
30
31 shocks have permanent effects on the distribution. In this case, the series does not
32
33 return to its former path, but follows a random growth path, under randomly
34
35 distributed shocks, as predicted by random growth theories. On the other hand, if
36
37 the series does not contain a unit root, the underlying process is deterministic as the
38
39 series has a short memory. Exogenous shocks do not have permanent effects and the
40
41 distribution returns to its steady state, after a certain time (DARNE and DIEBOLT,
42
43 2005).
44
45
46
47
48
49
50
51
52
53

54 The Chinese panel comprises 220 cities, and the Indian one 222 cities. For each of
55
56 these cities, annual data is available from the first to the last year of the reference
57
58 period. We run two first generation panel unit root tests, the LEVIN and al (2002)
59
60

1
2
3 and the IM and al (2003) tests for both specifications (with or without trend). Table 10
4
5
6 provides the results.
7
8
9

10 Table 10

11
12
13 The IM and al (2003) test rejects the H_0 hypothesis of non stationarity for Indian
14 cities for the model without trend. This means that there is at least one city for which
15 size is stationary. The non stationarity hypothesis cannot be rejected for Chinese
16 cities. In contrast, the LEVIN and al (2002) test rejects H_0 for both countries, which
17 goes against the results of the previous test. According to HURLIN and MIGNON
18 (2005), comparing the results of the LL and the \bar{Z}_t or the \bar{W}_t statistics may be
19 spurious. This leads to abandoning the restrictive assumption of cross section
20 dependence, by admitting that city size growth processes may be correlated. We use
21 the CHOI (2002) and PESARAN (2007) second generation panel unit root tests. Table
22 11 provides the results for these tests.
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42

43 Table 11

44
45
46 Both tests reject the null hypothesis of non stationarity for Chinese and Indian
47 city sizes. For China, they confirm the results of the IM and al (2003) tests. For India,
48 they prove that it is necessary to take into account the existence of cross-section
49 dependence in order to reach a clear conclusion. This means, firstly, that Gibrat's law
50 for cities cannot be entirely rejected in both countries. Secondly, in India, the urban
51
52
53
54
55
56
57
58
59
60

1
2
3 growth processes may be correlated. Thirdly, in a more general way, the results of
4
5 previous empirical studies based upon a single first generation panel unit roots test
6
7 may lead to an over-rejection of non stationarity for city sizes and the Gibrat's law.
8
9

10
11
12 Panel unit root tests do not allow one to conclude whether any shifts occurred to
13
14 the cities' positions within the rank size distribution. We use the Markov chains in
15
16 order to model relative urban growth and intra-distributional mobility for cities.
17
18 Tables 12 and 13 deliver the results of the discretisation process for the Chinese and
19
20 the Indian city size distributions. We proceeded to model the transition probabilities
21
22 which represent the number of cities moving from one cell to another every year.
23
24 Following BLACK and HENDERSON's (2003) work, we also calculate mean first
25
26 passage times.
27
28
29
30
31
32
33

34
35
36 Table 12
37

38
39
40 Table 13
41

42
43 The transition probabilities patterns for the Chinese cities (table 14) are driven by
44
45 two features: firstly, diagonal elements, that is, the probability of remaining in the
46
47 starting state is logically high, but off-diagonal elements are significantly different
48
49 than 0. In low city size groups, upward mobility is higher than downward mobility,
50
51 while it is the opposite for higher sizes. This is due to the anti-megacities policies and
52
53 the effects of the migration restrictions until the mid-nineties. The mean first passage
54
55
56
57
58
59
60

1
2
3 time matrix (table 15) helps characterise these results. It takes 21 years for a city from
4
5
6 group c_1 to visit group c_2 but 364 years for a city from group c_4 to visit group c_5 .
7
8
9

10 Table 14
11

12
13 Table 15
14
15

16
17 We proceeded to model transition probabilities (table 16) and mean first passage
18 time (table 17) also for Indian cities. Firstly, diagonal elements are higher than those
19
20 of the Chinese matrix. This indicates a higher stability in Indian urban hierarchies.
21
22
23 Secondly, unlike China, downward mobility is higher than upward mobility, except
24
25 for the larger cities. The mean first passage time matrix clearly shows that upward
26
27 mobility for cities is slower in India than in China: it would take 1707 years for an
28
29 Indian city from the c_1 group to visit group c_5 instead of 587 years for a Chinese city.
30
31
32 Higher volatility in Chinese urban hierarchies might be due to the fact that in China,
33
34 urban growth sources rely upon regional migration while Indian demographics
35
36 depend mainly on the natural increase in population (BHAGAT, 2005, DATTA,
37
38 2006).
39
40
41
42
43
44
45
46
47
48
49

50 Table 16
51

52
53 Table 17
54
55

56
57 When modeling the stationary (ergodic) state, it is clear that there is no
58
59 convergence process towards a single city size, for both distributions. Despite anti-
60

1
2
3 megacity policies in China, urban growth patterns lead to a higher instability in the
4
5
6 country's city size distribution, which would match with the predictions of random
7
8
9 growth theories. In India, the Markov chains feature rather parallel growth patterns
10
11
12 with small cities and large metropolises essentially remaining within their initial
13
14
15 state group.

16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60

CONCLUSION

This paper's findings can be summarized as follows:

Firstly, by using the Kolmogorov-Smirnov test, we show that the Chinese and Indian city-size distributions for cities of over 100,000 inhabitants follow a Pareto law. This contradicts recent findings from some authors who argue that the city-size distribution is log-normal. The definition of a city here plays a crucial role.

Secondly, the evolution of these two distributions over the last twenty years differs. The Chinese rank-size distribution reveals contrasting evolution between the 1980s, when small-size cities grew faster, and the 1990s, with higher growth trends for medium-size cities. The Indian city-size distribution systematically rejects Zipf's law, with a Pareto coefficient higher than 1. It is characterised by a dual structure with a substantial presence of small cities and a small group of very large metropolises.

1
2
3
4 Thirdly, when it comes to urban growth patterns, city size is not stationary in
5
6 both countries. Gibrat's principle cannot be rejected. Intra-distributional mobility is
7
8 higher in China than in India despite previous cross-region migration restrictions
9
10 and anti-megacity policies. Indian urban hierarchies feature a much higher stability
11
12 which could reveal parallel growth patterns.
13
14
15

16
17
18 One of the limitations of such a study is the short 20 year period for measuring
19
20 major shifts in city-size distribution. This time limitation was imposed because of
21
22 data availability. However, the results deliver some interesting comparative
23
24 conclusions on city-size dynamics in the two most populated countries in the world,
25
26 which went through significant economic and political shifts during this period.
27
28
29
30
31
32

33 34 NOTES

- 35
36
37
38
39 1. As is widely documented in the literature, the Hukou system (residence
40
41 permit) has been a critical barrier to rural-urban and inter-city migration in
42
43 China, leading to a spatially segmented labour market. In 1993, a series of
44
45 political reforms promoted both rural to urban and cross-region migrations,
46
47 except for the most important metropolitan areas, which still remain under the
48
49 control of anti-megacity policies.
50
51
52
53
54
55
56 2. According to GABAIX and IOANNIDES (2004), most comparative studies
57
58 tend to reject rather too often the assumption of a quadratic term equal to 0.
59
60

1
2
3 This is due to the bias of the OLS method, when applied to small samples,
4
5
6 which tends to underestimate the true standard error.
7
8
9

- 10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
3. For India annual city sizes are not provided by the Official Census. We have created these series using two different databases: those on natural increases in urban population, annually published by the Registrar General of India, and those on annual rural-urban migration, available from the *Migration tables*, published by India's Department of Statistics, Planning and Public Grievance. When comparing the Census data of 1991 and 2001 with the updated tables for cities' population of the 1981 Census, the results are fairly close (although generally slightly higher in 1991, with some rural-urban migration probably failing official registration).
4. According to the National Bureau of Statistics of China, urban and rural populations are classified in accordance with their usual residence. The urban population includes all people administered at city or town level. Towns include areas of more than 3000 inhabitants, with a non agricultural population of over 85%. This leads to an ambiguous definition of the total Chinese urban population, including people from very small districts.
5. The medial is the quantity such that the sum of all data values inferior to the medial equals the sum of all the superior data values.

REFERENCES

1
2
3
4
5
6
7
8 ANDERSON G. and GE Y. (2005) The size distribution of Chinese cities, *Regional*
9
10 *Science and Urban Economics*, **35**, pp.756-776.

11
12
13 BHAGAT R. (2005) Urban growth by city and town-size in India, *Acts of the*
14
15 *Population Association of America Conference*, Philadelphia, March 2005.

16
17
18 BLACK D. and HENDERSON J.V. (1999) A Theory of Urban Growth, *Journal of*
19
20 *Political Economy*, **107**, pp.252-284.

21
22
23 BLACK D. and HENDERSON J.V. (2003) Urban evolution in the USA, *Journal of*
24
25 *Economic Geography*, **3**, pp.343-372.

26
27
28 BOSKER E., BRAKMAN D., GARRETSEN H. and SCHRAMM M. (2008) A
29
30 century of shocks: the evolution of the German city-size distribution, *Regional Science*
31
32 *and Urban Economics*, **38-4**, pp.330-347.

33
34
35 CAMERON T. A. (1990) One-Stage Structural Models to Explain City Size, *Journal*
36
37 *of Urban Economics*, **27**, pp.294-307.

38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
CHOI I. (2002) *Combination Unit root tests for cross-sectionally correlated panels*,
Mimeo, Hong Kong University of Science and Technology.

CORDOBA J.C. (2008) On the Distribution of City Sizes, *Journal of Urban*
Economics, **63 (1)**, pp.177-197.

CUBERES D. (2004) The rise and decline of cities, *Working Paper*, University of
Chicago.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

DARNE O. and DIEBOLT C. (2005) Non-stationary tests in Macroeconomic Time Series, in DIEBOLT C. and KYRTSOU C. (2005) *New trends in macroeconomics*, Springer, Berlin.

DATTA P. (2006) Urbanization in India, *working paper*, Indian Statistical Institute, Kolkata.

DAVIS D.R. and WEINSTEIN D.E. (2002) Bones, Bombs and Breakpoint: the Geography of Economic Activity, *American Economic Review*, **92**, pp. 1269-1289.

DIMOU M. and SCHAFFAR A. (2007) Evolution des hiérarchies urbaines et loi de Zipf. Le cas des Balkans, *Région et Développement*, **25**, pp.65-86.

DIMOU M., SCHAFFAR A., CHEN Z. and FU S., 2008, La croissance urbaine chinoise reconsidérée, *Région et Développement*, **26**, pp.109-131.

DIMOU M. and SCHAFFAR A. (2009) Urban Hierarchies and city-growth in the Balkans, *Urban Studies*, **46(13)**, pp.2891-2906.

DELGADO A.P. and GODINHO I. (2007) Growth and Change in the Portuguese urban system: 1890-1901, *Acts of the International Conference on Regional and Urban modelling Conference*, Brussels, 1-2 June.

DOBKINS L.H. and IOANNIDES Y.M. (2000) Dynamic evolution of U.S. cities, in HURIOT J. and THISSE J (Eds.), *The Economics of Cities, Theoretical Perspectives*, Cambridge University Press, Cambridge, pp. 217-260.

DURANTON G. (2006) Some foundations for Zipf's law: product proliferation and local spillovers, *Regional Science and Urban Economics*, **36**, pp.542-563.

1
2
3 DURANTON G. (2007) Urban evolutions: the fast, the slow and the still, *The*
4
5
6 *American Economic Review*, **97(1)**, pp.197-221.
7

8
9 EATON J. and ECKSTEIN Z. (1997) Cities and Growth: Theory and Evidence
10
11 from France and Japan, *Regional Science and Urban Economics*, **XXVII**, pp. 443-474.
12

13
14 EECKHOUT J. (2004) Gibrat's Law for (all) Cities, *The American Economic Review*,
15
16
17 **94(5)**, pp.1429-1451.
18

19
20 FUJITA M., KRUGMAN P. and VENABLES A. (1999) *The Spatial Economy*, MIT
21
22 Press, Cambridge, MA.
23

24
25 GABAIX X. (1999) Zipf's Law for Cities: an Explanation, *Quarterly Journal of*
26
27 *Economics*, **114**, pp.739-767.
28

29
30
31 GABAIX X. and IBRAGIMOV R. (2009) RANK-1/2: A simple way to improve the
32
33 OLS estimation of tail exponents, *Journal of Business & Economic Statistics*,
34
35
36 forthcoming.
37

38
39 GABAIX X. and IOANNIDES Y. (2004) The evolution of city sizes' distribution in
40
41 HENDERSON J.V and THISSE J-F. (eds) *Handbook of regional and urban economics*,
42
43 vol.4, Elsevier Science B.B, Amsterdam, pp.2341-2376.
44

45
46
47 GAN L., LI D. and SONG S. (2006) Is the Zipf law spurious in explaining city-size
48
49 distributions? *Economic Letters*, **92**, pp.256-262.
50

51
52
53 GARMESTANI A., ALLEN C., GALLAGHER C. and MITTELSTAEDT J. (2007)
54
55 Departures from Gibrat's Law, discontinuities and city size distributions, *Urban*
56
57 *Studies*, **44**, pp.1997-2007.
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

GLAESER E., SCHEINKMAN J. and SHLEIFER A. (1995) Economic growth in a cross-section of cities, *Journal of Monetary Economics*, **36**, pp.117-143.

GUÉRIN-PACE F. (1995) Rank size distributions and the process of urban growth, *Urban Studies*, **32(3)**, pp.551-562.

HENDERSON V. and WANG H. (2007) Aspects of the rural-urban transformation of countries, *Journal of Economic Geography*, **5(1)**, pp.23-42.

HSING Y. (1990) A Note on Functional Forms and the Urban Size Distribution, *Journal of Urban Economics*, **27**, pp.73-79.

HURLIN C. and MIGNON V. (2005) Une synthèse des tests de racine unitaire sur données de panel, *Economie et Prévision*, **169**, pp.253-294.

IM, K., PESARAN, M. and SHIN, Y. (2003) Testing for Unit Roots in Heterogeneous Panels, *Journal of Econometrics*, **115**, pp.53-74.

IOANNIDES Y.M. and OVERMAN H.G. (2003) Zipf's Law for Cities: An Empirical Examination, *Regional Science and Urban Economics*, **33**, pp.127-137.

KRATZ M. and RESNICK S. (1996) The QQ-estimator and heavy tails, *Communications in Statistics, Stochastic models*, **12**, pp.699-724

KRUGMAN P. (1996) Confronting the mystery of Urban Hierarchy, *Journal of the Japanese and the International Economies*, **10**, pp.399-418.

LEVIN A., LIN C.F. and CHU C.S. (2002) Unit root tests in panel data: asymptotic and finite-sample properties. *Journal of Econometrics*, **108**, pp.1-24.

1
2
3 MALECKI E. (1980) Growth and change in the analysis of rank–size distributions:
4 empirical findings, *Environment and Planning A*, **12**, pp.41–52.
5
6

7
8 MORICONI-EBRARD F. (1993) *L'urbanisation du monde depuis 1950*, Paris,
9 Anthropolos.
10
11

12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

NOTA S. and SONG F. (2006) Further Analysis of the Zipf's Law: Does the Rank-Size Rule Really Exist? UNR Joint Economics Working Paper Series, *Working Paper*, 07-2006, University of Nevada.

PARR J. (1985) A Note on the Size Distribution of Cities Over Time, *Journal of Urban Economics*, **18**, pp.199-212.

PESARAN H. (2007) A simple panel unit root test in the presence of cross-section dependence, *Journal of Applied Econometrics*, **22(2)**, pp.265-312.

QUAH D. (1993) Empirical cross-section dynamics in economic growth, *European Economic Review*, **37**, pp.426-434.

REED W. (2001) The Pareto, Zipf and other power laws, *Economics Letters*, pp.7415-19.

ROSEN K. and RESNICK M. 1980, The size distribution of cities: an examination of the Pareto law primacy, *Journal of Urban Economics*, **8**, pp.165-186.

ROSSI-HANSBERG E. and WRIGHT M. (2007) Urban Structure and Growth, *Review of Economic Studies*, **74**, pp.597-624.

SHARMA S. (2003) Persistence and stability in city growth, *Journal of Urban Economics*, **53**, pp.300-320.

1
2
3 SIMON H. (1955) On a class of skew distribution functions, *Biometrika*, **44**, 425-
4
5
6 440.

7
8
9 SONG S. and ZHANG K. (2002) Urbanization and City Size Distribution in
10
11
12 China, *Urban Studies*, **39**, pp.2317-2327.

13
14 SOO K.T. (2005) Zipf's Law for cities: a cross-country investigation, *Regional*
15
16
17 *Science and Urban Economics*, **35**, pp.239-263.

18
19
20 XU Z. (2008) *Urbanisation et croissance des villes en Chine*, Ph.D Thesis, University
21
22
23 of Clermont-Ferrand.

24
25 ZEYNELOGLU S., KUNDAK S. and DOKMECI V. (2005) Methods and Data
26
27
28 consideration related to the rank size distributions of settlements, *Acts of the ERSA*,
29
30
31 Amsterdam, August.

32
33
34 ZHOU Y. and MA J. (2003) China's Urbanization Levels: Reconstructing a
35
36
37 Baseline from the Fifth Population Census, *The China Quarterly*, **173**, pp.176-196.

38 39 40 41 DATABASES

42
43
44
45 China Urban Statistical Yearbooks.

46
47
48 Indian Census.

49
50
51 Department of Statistics, Planning and Public Grievance, India.
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

TABLES

(to be inserted in the text)

For Peer Review Only

Table 1: Sample of Chinese cities 1984 - 1994 - 2004

Year	Number of cities n	Mean City size m	Standard deviation σ	Median city size Q ₅₀	First quartile Q ₂₅	Third quartile Q ₇₅	Medial ⁵ City size m _L	Rank for Medial city R _m
1984	225	467.6	731.4	245.6	150.7	446.9	880.5	29
1994	472	384.3	647.6	212.0	147.6	365.4	543.1	64
2004	577	468.5	818.0	239.0	155.0	436.6	755.7	72

Numbers in thousands of inhabitants.

Database: Chinese Urban Statistical Yearbooks.

Table 2: Sample of Indian cities 1981 - 1991 – 2001

Year	Number of cities N	Mean City size m	Standard deviation S	Median city size Q ₅₀	First quartile Q ₂₅	Third quartile Q ₇₅	Medial City size m _L	Rank for Medial city R _m
1981	225	361.5	742.3	170.2	120.1	305.8	606.3	24
1991	313	381.5	821.7	172.7	125.0	324.9	690.8	32
2001	422	402.9	904.5	177.0	126.3	315.5	847.1	41

Numbers in thousands of inhabitants.

Database: Census of India.

Table 3: Results of the KS-test for Chinese cities (Pareto)

Year	Number of cities	Pareto exponent	KS-stat	p-value
1984	225	0.912	0.901	0.391
1994	472	0.993	1.093	0.183
2004	577	0.884	1.212	0.106

Null hypothesis is rejected at 5% (*), at 1% (**)

* Pareto exponent for a minimum p-value

Table 4: Results of the KS-test for Indian cities (Pareto)

Year	Number of cities	Pareto exponent	KS-stat	p-value
1981	225	1.356	0.696	0.717
1991	313	1.247	0.554	0.918
2001	422	1.179	0.393	0.998

Null hypothesis is rejected at 5% (*), at 1% (**)

* Pareto exponent for a minimum p-value

Table 5: Results of the KS-test for Chinese cities (Log-normal)

Parameters for the log-normal Distribution	Number of cities <i>n</i>	Mean Size <i>m</i>	Standard deviation σ	KS-stat	p-value
Sample	577	5.67	0.82	2.32 (**)	0
Anderson and Ge	661	4.20	0.70	14.22 (**)	0
Districts & non-agricultural pop.	860	4.17	1.1	3.53 (**)	0

Null hypothesis is rejected at 5% (*), at 1% (**)

Databases: CSY (2005), Anderson and Ge (2005)

Table 6: Results of the KS-test for Indian cities (Log-normal)

Parameters for the log-normal Distribution	Number of cities n	Mean Size m	Standard deviation σ	KS-stat	p-value
Sample	422	5.46	0.82	3.62 (**)	0
Districts, Total urban Population	5161	3.29	1.04	3.59 (**)	0

Database: Census 2001

Table 7: Pareto exponent for the Chinese and Indian city-size distribution

Year	Sample's size n	Pareto exponent β	Standard deviation $\sigma(\beta)$	Student t	p-value α	Decision $t \neq 1$
China						
1984	225	1.076	0.073	1.422	15.06%	No
1994	472	1.198	0.059	2.745	3.01%	Yes
2004	577	1.089	0.053	1.656	19.20%	No
India						
1981	225	1.256	0.118	2.159	3.23%	Yes
1991	313	1.226	0.098	2.308	2.28%	Yes
2001	427	1.203	0.082	2.468	1.45%	Yes

Table 8: Sensitivity of the estimated parameters for the Chinese rank size distribution
according to the city-size threshold

City-size threshold S_{\min}	Year	Number of cities n	Pareto exponent β	Standard deviation $\sigma(\beta)$	Student t
100,000	1984	225	1.076	0.073	1.422
	1994	472	1.198	0.059	2.745(*)
	2004	577	1.089	0.053	1.656
200,000	1984	130	1.221	0.164	1.959
	1994	250	1.320	0.127	3.306(**)
	2004	352	1.242	0.101	3.379(**)
500,000	1984	49	1.410	0.167	2.080(*)
	1994	72	1.446	0.238	2.120(*)
	2004	129	1.493	0.302	2.722(**)
1000,000	1984	19	1.684	0.546	1.462(*)
	1994	32	1.755	0.439	1.721(**)
	2004	49	1.690	0.341	2.020(*)

Null hypothesis ($\beta=1$) is rejected at 5% (*), at 1% (**)

Table 9: Sensitivity of the estimated parameters for the Indian rank size distribution
according to the city-size threshold

City-size threshold S_{\min}	Year	Number of cities n	Pareto exponent β	Standard deviation $\sigma(\beta)$	Student t
100,000	1981	225	1.284	0.083	3.427 (**)
	1991	313	1.233	0.070	3.347 (**)
	2001	427	1.177	0.060	2.973 (**)
200,000	1981	99	1.324	0.129	2.517 (*)
	1991	129	1.161	0.109	1.467
	2001	210	1.139	0.079	1.765
500,000	1981	29	1.282	0.198	1.421
	1991	49	1.395	0.195	2.027 (*)
	2001	68	1.295	0.195	1.901
1,000,000	1981	11	1.072	0.270	0.266
	1991	18	1.303	0.394	0.768
	2001	27	1.348	0.327	1.063

Null hypothesis ($\beta=1$) is rejected at 5% (*), at 1% (**)

Table 10: Results of the first generation panel unit root tests – China and India

	Test	Statistics	Model with trend	Model without trend	
China	Levin, Lin and Chu (2002)	LL	-3,916	-4,114	
			(0,00)	(0,00)	
	Im, Pesaran and Shin (2003)	\bar{Z}_t	-4,835	18,182	
			(0,00)	(1,00)	
			\bar{W}_t	4,472	19,975
			(1,00)	(1,00)	
India	Levin, Lin and Chu (2002)	LL	-6,696	-8,556	
			(0,00)	(0,00)	
	Im, Pesaran and Shin (2003)	\bar{Z}_t	-9,960	-15,707	
			(0,00)	(0,00)	
			\bar{W}_t	0,260	-14,977
			(0,60)	(0,00)	

(p values within brackets)

Table 11: Results of the second generation panel unit root tests – China and India

	Test	Statistics	Model with trend	Model without trend
China	Choi (2002)	P_m	-1,007	-10,476
			(0,84)	(1,00)
		Z	7,534	14,615
	Pesaran (2007)	L^*	9,307	10,005
			(1,00)	(1,00)
		$CIPS$	-1,347	-1,474
India	Choi (2002)	P_m	-7,987	-11,671
			(1,00)	(1,00)
		Z	10,179	9,110
	Pesaran (2007)	L^*	10,591	8,598
			(1,00)	(1,00)
		$CIPS$	-2,194	-1,890
		$CIPS^*$	-2,194	-1,890
			(0,71)	(0,23)
			(0,71)	(0,23)

(p-values within brackets)

Table 12: Chinese discretised city size distribution

Groups	City Size (1000 inhs)	Number of cities	Share in %
$c_1 : S_{\min} \leq S \leq 0,3m$	100 - 140,3	47	21,4
$c_2 : 0,3m < S \leq 0,5m$	140,3 - 233,8	60	27,3
$c_3 : 0,5m < S \leq m$	233,8 - 467,6	61	27,7
$c_4 : m < S \leq 2m$	467,6 - 935,2	28	12,7
$c_5 : 2m \leq S$	$S > 935,2$	24	10,9

Review Only

Table 13: Indian discretised city size distribution

Groups	City Size (1000 inhs)	Number of cities	Share in %
$c_1 : S_{\min} \leq S \leq 0,3m$	100 - 108,5	25	11,2
$c_2 : 0,3m < S \leq 0,5m$	108,5 - 180,8	92	41,3
$c_3 : 0,5m < S \leq m$	180,8 - 361,5	62	27,7
$c_4 : m < S \leq 2m$	361,5 - 723	26	11,7
$c_5 : 2m \leq S$	$S > 723$	18	8,1

Table 14: Transition matrix- China

\hat{p}_{ij}	c_1	c_2	c_3	c_4	c_5
c_1	0,952 (0,0084)	0,048 (0,0084)	0	0	0
c_2	0,014 (0,0032)	0,960 (0,0054)	0,026 (0,0044)	0	0
c_3	0	0,015 (0,0032)	0,974 (0,0043)	0,011 (0,0028)	0
c_4	0	0	0,024 (0,0060)	0,964 (0,0074)	0,013 (0,0044)
c_5	0	0	0	0,026 (0,0078)	0,974 (0,0078)

Table 15: Mean first passage time - China

$M_{p,ij}$	c_1	c_2	c_3	c_4	c_5
c_1	0	20,8	70,0	222,2	586,7
c_2	288,3	0	49,2	201,4	565,9
c_3	402,8	114,5	0	152,2	516,7
c_4	465,5	177,2	62,7	0	364,5
c_5	503,7	215,4	100,9	38,27	0

Table 16: Transition matrix- India

\hat{p}_{ij}	c_1	c_2	c_3	c_4	c_5
c_1	0,985 (0,0041)	0,015 (0,0041)	0	0	0
c_2	0,028 (0,0042)	0,964 (0,0047)	0,008 (0,0023)	0	0
c_3	0	0,019 (0,0042)	0,966 (0,0055)	0,015 (0,0037)	0
c_4	0	0	0,011 (0,0043)	0,981 (0,0055)	0,008 (0,0036)
c_5	0	0	0	0,011 (0,0056)	0,989 (0,0056)

Table 17: Mean first passage time - India

$M_{p,ij}$	c_1	c_2	c_3	c_4	c_5
c_1	0	66,5	405,5	896,3	1707,0
c_2	89,0	0	339,1	829,8	1640,5
c_3	264,2	175,2	0	490,8	1301,4
c_4	416,2	327,2	152,0	0	810,7
c_5	505,5	416,5	241,3	89,3	0

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

FIGURES

(to be inserted in the text)

For Peer Review Only

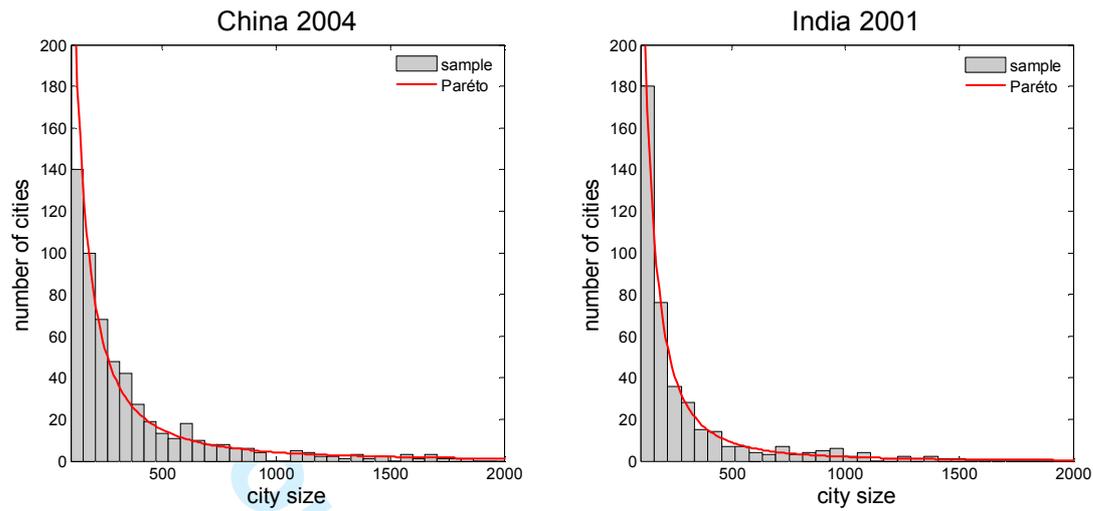


Figure 1: Histograms of Chinese and Indian cities with sizes above 100,000 inhabitants in 2004 and 2001

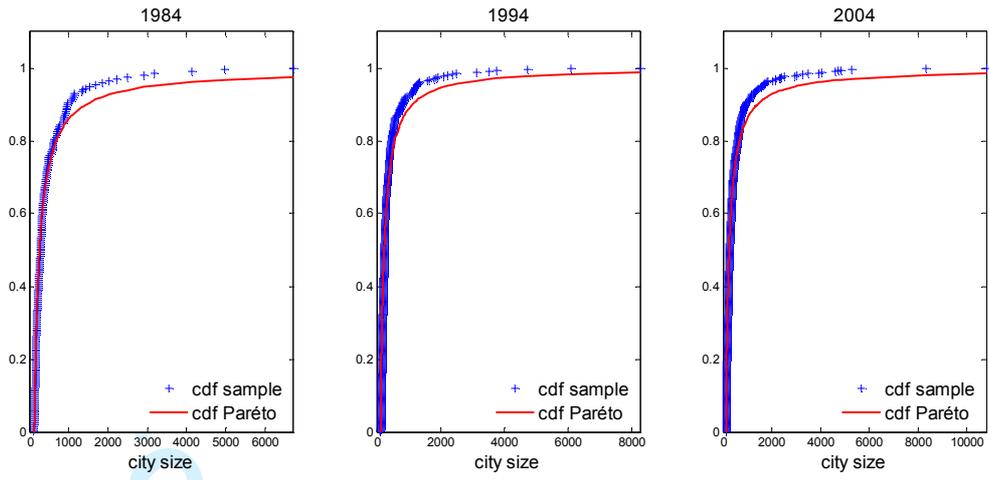


Figure 2: Distribution function for Chinese cities

with more than 100,000 inhabitants

Review Only

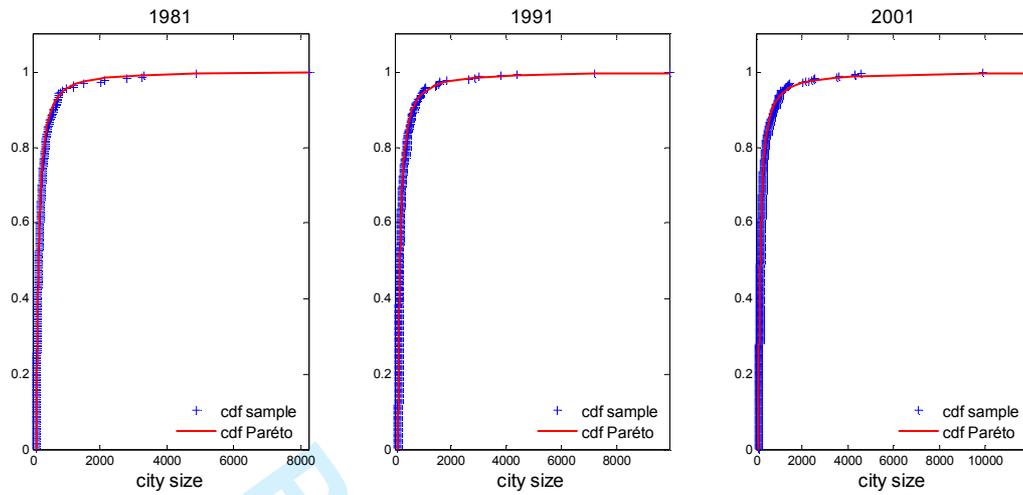


Figure 3: Distribution function for Indian cities

with more than 100,000 inhabitants

view Only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

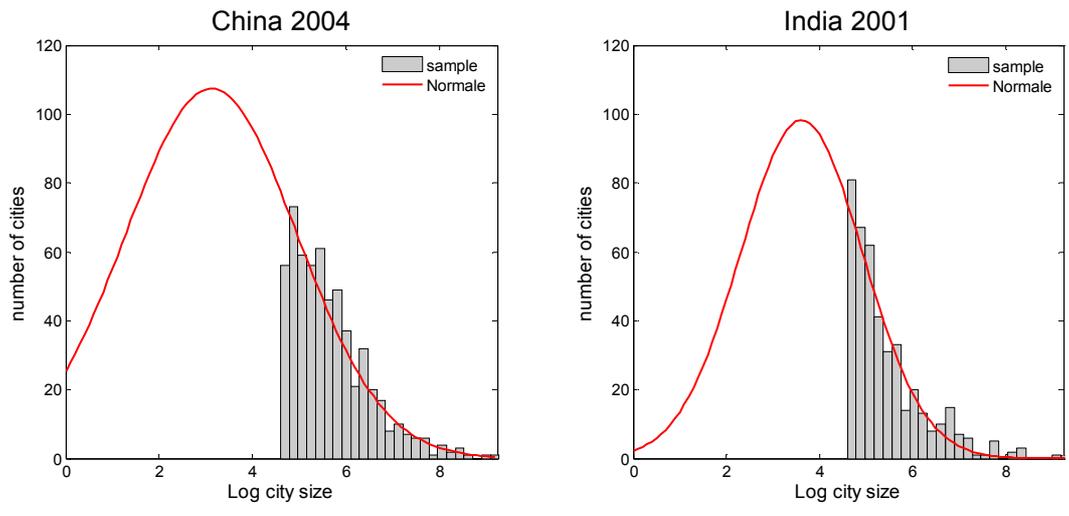


Figure 4: Hypothesis of a truncated log-normal distribution

er Review Only

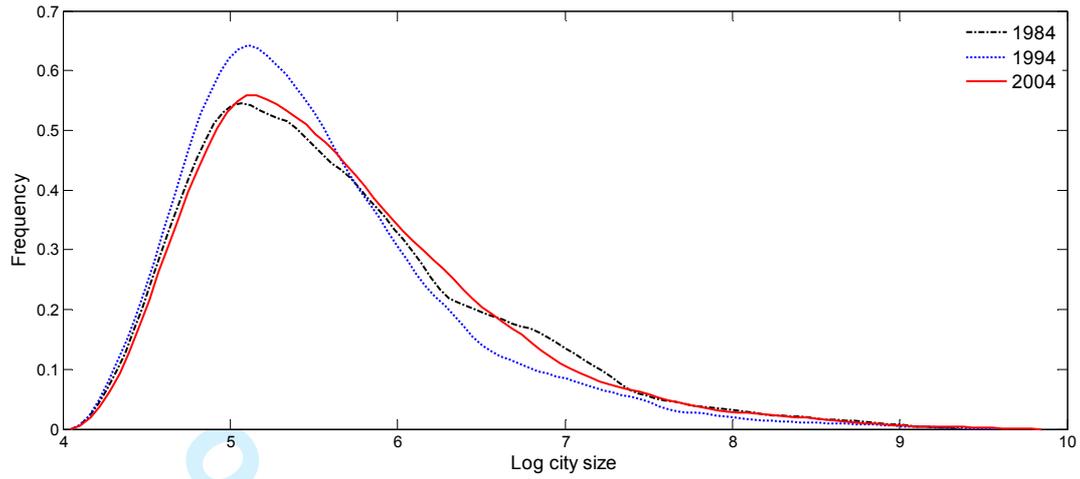


Figure 5: Density kernels for Chinese city-sizes 1984-1994-2004

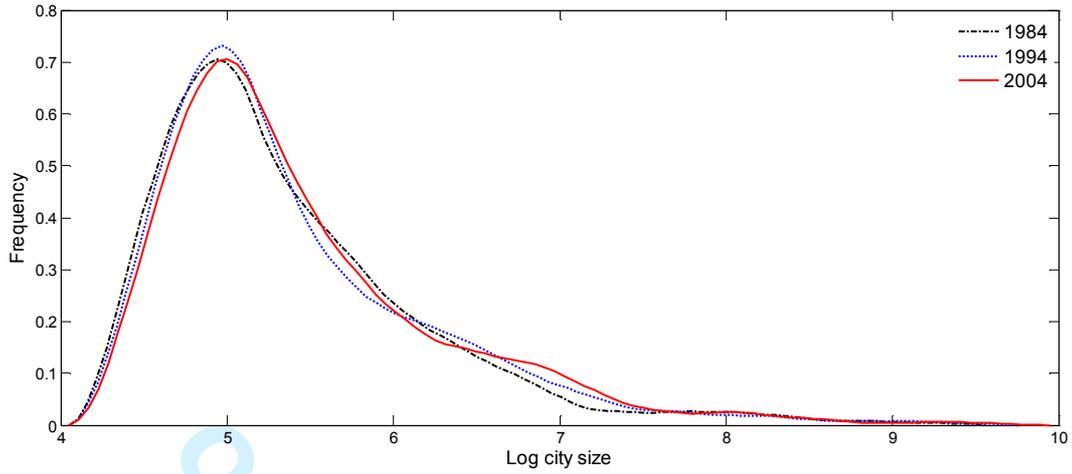


Figure 6: Density kernel for Indian cities 1981-1991-2001

Peer Review Only