



Stanford
Center for International
Development

Working Paper No. 415

Zipf's Law for Chinese Cities: Rolling Sample Regressions

by

Guohua Peng

March 2010



Stanford University

John A. and Cynthia Fry Gunn Building, 366 Galvez Street

Stanford, CA 94305-6015

Zipf's Law for Chinese Cities: Rolling Sample Regressions

Guohua Peng*

College of Economics, Jinan University, Guangzhou, 510632, People's Republic of China

Abstract

We study the validity of Zipf's Law in a data set of Chinese city sizes. Previous investigations are restricted to log-log rank-size regression for a fixed sample. In contrast, we use rolling sample regression methods in which the sample is changing with the truncation point. The intuition is that if the distribution is Pareto with a coefficient one (Zipf's law holds), rolling sample regressions should yield a constant coefficient regardless of what the sample is. We find that the Pareto exponent is almost monotonically decreasing in the truncation point; the mean estimated coefficient is 0.84 for the full dataset, which is not so far from 1.

Keywords: city size; Zipf's law; Pareto distribution; rolling sample regressions

JEL classification: C16; R11;R12

*At present, the author is a visiting researcher at Stanford University. The present address: John A. and Cynthia Fry Gunn Building, 366 Galvez Street, Stanford, CA 94305-6015, USA. Phone: (650)724-5138, Fax: (650)725-6069. Email: pengguohua@gmail.com.

1 Introduction

The well-known regularity of Zipf's law concerning city size distribution¹ which refined from Auerbach [1], has been widely investigated across countries and over time [2-4]. In the upper tail of city size distribution, a Pareto distribution could be a good approximation to the data; and, Zipf's law appears when the shape parameter of the distribution is equal to one [5].² Many papers estimate the Pareto exponent to test the Zipf's law. When the value of the Pareto exponent is not different from one, it is said that the Zipf's law holds; otherwise, Zipf's law is rejected. Rosen and Resnick [3] find that the exponent in 32 out of 44 countries exceeded one with a sample mean of 1.14 which is close to 1. Using a new and larger cross-country dataset, Soo [4] rejects the Zipf's law for 53 out of 73 countries using ordinary least squares (OLS). Soo [4] finds an average Pareto exponent of 1.11, it is pretty close to Zipf's law, though it is not exactly 1. Most previous investigations are restricted to log-log rank-size regression for a fixed sample. It rules out the probability of variation for the exponent by changing the sample size. Some authors including Rosen and Resnick [3], Black and Henderson [9], and Eeckhout [10] suggest that the value of the exponent is sensitive to the choice of sample size. However, there is no empirical paper that systematically investigates it.

In this paper, we follow a different approach; by using rolling sample regression methods rather than fixed sample size, we find some deviations from Zipf's law in Chinese city size data. The intuition is that if the distribution is Pareto with a coefficient one (Zipf's law holds), rolling sample regressions should yield a constant coefficient independent of the sub-sample size for the rolling regressions.

The Pareto distribution with shape parameter has the counter-cumulative distribution function: $1 - F(x) = P[S > x] = (a/x)^\beta$, $a, x > 0, x \geq a$, with the corresponding probability density function: $p(x) = \beta a^\beta x^{-\beta-1}$. The Pareto distribution is the only distribution that is invariant under truncation from below, which means that for the Pareto distribution, for $y \geq S^*$, $P[S > y | S > S^*] = P[(S^*/a)S > y] = (S^*/y)^\beta$. So the conditional distribution is also a Pareto distribution, with the same shape parameter and new location parameter S^*

¹Zipf's law for cities states that a city of rank r in the (descending) order of cities has a size $S_{(r)}$ equal to $1/r$ times the size of the largest city in that country. That's to say, the size $S_{(2)}$ of the second largest city should be $1/2$ the size of the largest city; the size $S_{(3)}$ of the third largest city should be $1/3$ the size of the largest city; the size $S_{(10)}$ of the tenth largest city should be $1/10$ the size of the largest city, etc. For example, this law fits the first five largest cities in the United States quite well.

²To get clear on the concept, we follow Axtell [6], Gabaix [7], Gabaix and Ioannides [8], Rosen and Resnick [3], and Soo [4], and state that Pareto distribution is also known as a Power law distribution, and, a power law distribution with exponent (at least approximately) equal to 1 is called Zipf's law.

[7,11,12]. This remarkable property of the Pareto distribution means that the rolling rank regressions are regressions of the log transformation of this equation for decreasing values of S^* . If the distribution is Pareto, rolling sample regressions should yield a constant coefficient (which in the case of Zipf’s law is not different from one).

There are two advantages of rolling sample regression methods. First, they can systematically uncover the variation pattern of the exponent coefficient with the variation of truncation point. Some empirical papers investigate the value of the exponent using cross-country data. If the number of cities is different, or the lower city size threshold varies, then the large variation of the value of the exponent for different countries is expected. Second, the rolling sample regression methods adapt to the fact that the research interest extends to the entire distribution from the upper tail of the distribution. The census data are no longer restricted to large cities; and, the truncation point for city size data is decreasing significantly in some countries and will decrease dramatically in other countries. More and more cities are included in dataset from the census. For instance, using Census 2000 data including places even with only one person, Eeckhout [10] shows that the entire size distribution is lognormal.

We employ two complementary rolling sample regression methods. One method is rolling rank regressions in which the starting point of each sub-sample is fixed at the largest city and the truncation point moves down one city every time; so, the sub-sample size is increasing by one each time. The other method is rolling window regressions in which the size of each sub-sample of regression is fixed; that is, a fixed-width window with the start point and end point simultaneously moving down one city every time, i.e., the width of the window remains unchanged.

We perform the analysis using OLS for two rolling sample regression methods.³ As shown by Gabaix and Ibragimov [15], OLS approach is widely used in the Zipf regression because of “arguably the simplicity and robustness of this method”. The improved OLS estimation proposed by Gabaix and Ibragimov [15] is adopted here.

We find that the Pareto exponent is decreasing in the truncated point, except for the very upper tail of a full sample size of about 650 cities each year in 1999-2004 for the rolling

³We also use the Hill [13] estimator for these two rolling sample regression methods. The Hill estimator is the maximum likelihood estimator under the null of Pareto distribution. The Hill estimator is almost the same as the MLE estimator discussed in Newman [14]. However, as Gabaix [7], Gabaix and Ioannides [8], and Soo [4] point out, both methods (Hill and OLS) have pitfalls. But the OLS approaches are more robust than the Hill estimator of a tail index under deviations from Pareto laws [15]. Our results of the Hill estimator are similar to those of OLS; however, the deviations of the Hill estimator are much larger than those of OLS. Detailed results are available from the author upon request.

rank regressions. Our estimates range from 0.799 (full sample in 2002) to 1.852 (the 100 largest cities sub-sample in 2002), with a sample mean of 0.84 for the full sample size. The results are not so far from the Zipf’s law, although there are some deviations from the law.⁴ The variation pattern of the Pareto exponent is similar for the rolling window regressions except that there are some fluctuations in the lower half of the full sample size.

Following the introduction, this paper is organized into the following three sections: section 2 describes the data and the methods; section 3 presents the results of Zipf’s law; section 4 concludes the study.

2 Data and methods

2.1 Data

We study the validity of Zipf’s law using China’s recent city size data; it is very interesting because of its rapid urbanization with the largest population in the world. Although Chinese urbanization percent is steadily increasing in recent years, a striking feature of China is the low proportion of the population (about 30%) which resides in cities compared to that of developed countries (for instance 80% in the United States in 2000, [17]). We use a data set of the Chinese city population from 1999 to 2004 obtained from the Chinese Urban Statistical Yearbooks (State Statistical Bureau, 2000-2005). The Chinese Urban Statistical Yearbooks report three administrative levels of cities including province-level, prefecture-level, and county-level. There are two types of city population information in the Chinese Urban Statistical Yearbooks including: (1) Shiqu (urban) population, the population of an urban area; and, (2) Diqu population, the population of an urban area plus rural areas under the city’s jurisdiction. In China, a city is an urban area surrounded by rural areas. The structure of a city is just like a cell, with an urban area as the core of the cell. Rural areas are included in a city only because the rural places are under the city’s jurisdiction for administrative considerations. While the population of Diqu is the total population of a city consists of the urban population and rural population, the population of Shiqu is the population concentrated in the urban area of a city. That is to say, Shiqu is closer to true agglomeration than Diqu. Many researchers such as [3, 4, 8], consider that agglomeration would be a better definition as it typically includes surrounding suburbs where the workers of

⁴Since China was not recognized as a market economy, the deviations from Zipf’s law are expected, as Ades and Glaeser [16] analyze the causality of political factors to urban concentration. The author thanks one referee who pointed this out.

a city reside. However, as Au and Henderson [18] (and reference therein) describe that China strongly restricts rural-urban migration through the hukou system, which is similar to an internal passport system in China. In addition, most of the workers in China cannot afford their own cars and cannot reside in the surrounding suburbs with jobs in the urban area. Therefore, Shiqu (urban area of a city) more closely approximates an integrated economic unit in China. Thus we use Shiqu population, where most of the people of a city resides, as does Anderson and Ge [19]. Moura Jr. and Ribeiro [2] also exclude the rural population from their sample in the study of the Zipf law for cities in Brazil.

Our sample period begins from 1999 to 2004 because before 1999, the Chinese Urban Statistical Yearbooks did not report data of county-level cities. Another point to consider is that during this period there were very few new cities entering the system. In addition, the final reason is to provide new evidence for Zipf’s law due to Soo [4], who assesses the empirical validity of Zipf’s law using a new cross-country dataset. Table 1 reports the descriptive statistics of the Chinese city size. The number of cities reported in the Chinese Urban Statistical Yearbooks ranges from a low of 592 in 2000 to a high of 664 in 2001. The minimum city size is from 9,000 persons in 2003 to 16,000 persons in 2004. Furthermore, the largest city size rises from 11.3 million to 12.9 million—Shanghai, located in the Yangzi River delta in the east of China. The urbanization share is steadily increasing from 25.5% to 31.8%; that is, approximately 93.9 million more persons living in cities during these years.

2.2 Methods

Zipf’s law, also referred to as the rank size rule, is a deterministic rule that states that the population size $S_{(r)}$ of a city in the decreasingly ordered sequence of n cities with their population $S_{(1)} \cdots \geq S_{(r)} \geq \cdots \geq S_{(n)}$ is inversely proportional to the rank of the size of the city (r). An OLS regression of the log rank r on the log size $S_{(r)}$ is a way to estimate the Pareto exponent β :

$$\ln r = C - \beta \ln S_{(r)}, \quad (1)$$

If the estimated $\hat{\beta}$ is not significantly different from one that suggests that Zipf’s law holds.

It is well known that OLS estimated coefficients and the standard errors are strongly biased downward in small samples [8]. Fortunately, there is a remedy for this problem provided by [14]. They show that a shift of 0.5 for the rank is optimal and could cancel the bias up to a leading order. Furthermore, the standard error on the Pareto exponent is equal to $(2/n)^{0.5} \hat{\beta}$, where n is the corresponding sample size. Therefore, the regression (1)

becomes regression (2) below:

$$\ln(r - 0.5) = C - \beta \ln S_{(r)}. \quad (2)$$

Two rolling sample regression methods are used in this paper including rolling rank regressions and rolling window regressions. Using rolling rank regressions, we will repeat the estimation process using a diminishing lower truncation point, which is the starting point of each sub-sample that is the same and fixed at the largest city. Furthermore, the lower truncation point moves down by one city every time so that the sub-sample size increases by one each time. For example, the full sample size is 654 in 2004; these cities will be in descending order from the largest city, Shanghai, to the smallest city, Shaoshan (Hunan Province). The first sub-sample size of the regression (2) is \bar{n}_1 , the 20 largest cities for instance. The second sub-sample is $\bar{n}_2 = \bar{n}_1 + 1$, the 21 largest cities. Next, the third is $\bar{n}_3 = \bar{n}_2 + 1$, or 22 largest cities. This process continues until the last sub-sample is 654 cities, which is the full sample size. We will get 645 (=654-20+1) estimated exponent coefficients of $\hat{\beta}$ by repeating regressions 645 times with such a rolling rank regression method.

The following describes the alternative rolling window regression method as: we repeat the OLS estimation of equation (2) with a constant sub-sample size (fixed-width window) rather than increasing the sample size as shown above. With the start point and end point moving down one city each time, the fixed-width window regressions would roll from the start point (the largest city) to the end point (the smallest city) of the full sample size. The width of the window we choose is 100, i.e., the sub-sample size of these regressions remains unchanged and equal to 100, which is similar to the average number (102) of cross-country cities in Soo [4]. For instance, in 2004 the first window (sub-sample) of the fixed-width regression is from the largest city (Shanghai) to the 100th largest city (Yongzhou, Hunan Province). The second window (sub-sample) of fixed-width regression is from the second largest city (Beijing, the capital of China) to the 101th largest city (Jining, Shandong Province), and so on. The last window (sub-sample) is the smallest 100 cities from Hejian (Hebei Province) to Shaoshan. There are 555 (=654-100+1) windows (regressions) for the year of 2004 which has 654 cities. The rank of the top city in any sub-sample (or any regression window, whose width is constant and equal to 100 in the paper) is called rank 1 rather than its actual rank in the full sample, because every window is regarded as an independent regression due to the “remarkable property of the Pareto distribution” discussed in the Introduction. For instance, we call the first city in the 500th regression window, which has 100 cities from 500th to 599th in the full sample, rank 1 instead of rank 500.

In order to compare the theoretical Zipf's law to the empirical distribution, we provide a simple simulation. Take for instance a sample that is the 654 cities of the year 2004; we produce an exact rank size rule for 654 cities with the largest city being just the No.1 of 2004, which has a population of 12,891.3 thousand persons. That is, the simulation size series is $\bar{S}_{(1)} = 12891.3, \bar{S}_{(2)} = \frac{\bar{S}_{(1)}}{2}, \bar{S}_{(3)} = \frac{\bar{S}_{(1)}}{3}, \dots, \bar{S}_{(654)} = \frac{\bar{S}_{(1)}}{654}$. We use two rolling sample regression methods described above for the simulation size series to demonstrate how the theoretical exponent will vary.

3 Results

3.1 Rolling rank regressions

We estimate equation (2) using the rolling rank regression method with the beginning sub-sample size of 20, an arbitrary start number for each year separately, and get 644 (=663-20+1), 573, 645, 636, 631, and 635 exponent coefficients respectively. The average value of the Pareto exponent for all regressions is about 1.3, while the average value of the Pareto exponent for the full sample is roughly 0.84. Actually, they are not so far from 1, which implies that Zipf's law fits the size distribution of Chinese cities well at least approximately. Table 2 reports the components of the results. When the estimated sub-sample is the top 100 cities, the value of the Pareto exponent is about 1.8. In addition, 2002 is the year with the largest value of the Pareto exponent with a value of 1.852, and, the lowest value of 1.781 is obtained for 2003. With the rolling sample size increasing, the estimated Pareto exponent decreases until the truncation point reaches the end of the full sample, where the value of the Pareto exponent is about 0.84 (with the largest value of 0.876 for 1999, the lowest value of 0.799 for 2002). Figure 1 illustrates the relationship between the Pareto exponent with 95% confidence interval and the truncation point from the 20th largest city to the smallest city for each year. A first glance at the figure suggests that the Pareto exponent is negatively related to the sub-sample size. Actually, the value of the estimated $\hat{\beta}$ is monotonically decreasing in the truncation point providing that the sub-sample size exceeds about 100.

Standard errors (in parentheses) in table 2 are corrected by the method in [15]. The statistical significance of the Pareto exponent shows that $\hat{\beta}$ is significantly greater than one when the sub-sample size is less than 300; and, $\hat{\beta}$ is significantly less than one when the sub-sample size is the full sample size. There are two ranges illustrated in Figure 1 within the Pareto exponent in which is not significantly different from one at a 5% level.

The last column of Table 2 clearly reveals the result for the simulation size series. As we

expect, the Pareto exponent is not significantly different from one for every sub-sample size. The average value is 1.036. The estimates are almost constant unless the sub-sample is very small. This variation pattern confirms that when the distribution is Pareto, then the rolling regressions should yield a constant coefficient and that in the case of Zipf's law should not be different from one.

Figure 2 shows a plot of log rank against log population for the full sample size of each year and of simulation size series. The scatter plot that is apparently not in a line but concave is in accordance with the fact that the Pareto exponent varies with the sub-sample size. The concave plot suggests that the large cities are too small and the small cities are underrepresented; Rossi-Hansberg and Wright [17] emphasize this cross-country general feature. Compared to the concave empirical plot of 1999-2004, the plot of simulation size series is a perfect line with a slope of negative one. That is, it reflects an exact rank size rule.

In addition, our empirical distribution tests reject the hypothesis that the full sample size distribution is lognormal with the zero p-values for the adjusted statistics of the Lilliefors test and the Anderson-Darling test for each year. Gan, Li, and Song [20] reach the same results for the 1985 and the 1999 Chinese city size data.

3.2 Rolling window regressions

For the rolling window regressions for each year separately with the fixed-width window of 100, we get 564 (=663-100+1), 493, 565, 556, 551, and 555 exponent coefficients respectively. The decreasing variation pattern of the Pareto exponent is similar to the rolling rank regressions except that there are some fluctuations in the lower half of the full sample size. The average value of the exponent is about 0.75 which is much smaller than that (1.3) of rolling rank regressions.

Table 3 reports the components of the results. The first regression window, where the estimated sub-sample is the top 100 cities, gives the same result as the rolling rank regressions when its truncation point is the 100th largest city. With the regression window rolling down, the estimated Pareto exponent decreases dramatically. The value of the exponent is already reaching one at the 100th regression window where the starting point of the window is the 100th largest city and the ending point is the 199th largest city. After that, there are some fluctuations of the exponent below one and then it dramatically decreases again until the last regression window, where the value of the Pareto exponent is only about 0.11 (the largest value of 0.126 for 2000; the lowest value of 0.093 for 2003). Figure 3 illustrates

the relationship between the Pareto exponent with a 95% confidence interval and regression windows for each year. The figure suggests that the Pareto exponent negatively relates to the location of regression windows as a whole. Except for one range the Pareto exponent is significantly different from one for every year. An exception is 1999, which has three such ranges.

The rolling window regression result for the simulation size series is almost the same as the rolling rank regressions. The Pareto exponent differs insignificantly from one for every window and the confidence interval is almost parallel since the sub-sample size of every window regression is constant (see Figure 3). The average value of the exponent is 1.003, the maximum is 1.059, and the minimum is 1.008. This variation pattern also confirms that when the distribution is Pareto then the rolling regressions should yield a constant coefficient; and, in the case of Zipf's law, it should be one. That is to say, we should get the same result for whichever rolling sample regression method is used for an exact rank size rule.

3.3 Nonparametric analysis of the distribution of the Pareto exponent

We also perform a nonparametric analysis for the distribution of the Pareto exponent for the rolling sample regressions to show a complete description of how the values of the Pareto exponent are distributed, as does Soo [4]. We construct the efficient Epanechnikov kernel function using the optimal window width for the Pareto exponent for both the rolling rank and rolling window regressions. Figure 4 demonstrates the kernel function for the rolling rank regressions. The empirical distributions for 1999-2004 are bimodal with the left modes higher than the right modes. The left modes are around one while the right modes are about 1.7. Nevertheless, the distribution is clearly unimodal for the simulation size series with the mode at approximately 1.022.

Figure 5 shows the kernel function for the rolling window regressions. Now, the distributions are unimodal rather than bimodal. There are some fluctuations around the modes, which just reflect the fact that there are more fluctuations in the values of the Pareto exponent for the rolling window regressions. The kernel function for the rolling window regressions of the simulation size series is very similar to that of the rolling rank regressions. It is unimodal with the mode at approximately 1.013, which is slightly smaller than that of rolling rank regressions. If small sub-samples (from 20 to 100) are used for the rolling rank regressions, we observe a somewhat larger estimate of the exponent. Other than that,

these two kernel functions for the Pareto exponent of the simulation size series are almost identical, which confirms that the distribution of the Pareto exponent should be similar for the two rolling regressions.

4 Conclusion

Using rolling sample regression methods rather than fixed sample size, we study the empirical validity of Zipf's Law for Chinese cities. The intuition is that if the distribution is Pareto with a coefficient one (Zipf's law holds), rolling sample regressions should yield a constant coefficient not different from one. We employ the following two complementary rolling sample regression methods: (1) rolling rank regressions with the start point of each sub-sample fixed at the largest city and the truncation point moving down one city every time; and, (2) rolling window regressions with a fixed-width window rolling down from the largest city to the smallest city. We find that the Pareto exponent is almost monotonically decreasing in the truncated point for the rolling rank regressions; the mean estimated coefficient is 0.84 for the full dataset, which is not so far from 1. This implies that Zipf's law fits the size distribution of Chinese cities well at least approximately. The variation pattern of the Pareto exponent is similar for the rolling window regressions except that there are some fluctuations in the lower half of the full sample size.

Acknowledgements

The author would like to thank the Editor, anonymous referees, Richard Arnott, Xavier Gabaix, Nick Hope, and Hongliang Zhang for helpful suggestions and comments, and thank SCID Director, Nick Hope, for gracious hospitality in hosting the author's visit at Stanford. The author gratefully acknowledges the research support by the National Social Science Foundation of China (No. 07AJL009 and 08CJL018), the National Science Foundation of China (No. 70873051), the Program for New Century Excellent Talents in University, and the Jinan University Innovation Foundation. All errors are the author's own.

References

- [1] F. Auerbach, Petermanns Geographische Mitteilungen 59 (1913) 74-76.
- [2] N.J. Moura Jr., M.B. Ribeiro, Physica A, 367 (2006) 441-448.

- [3] K.T. Rosen, M. Resnick, *Journal of Urban Economics* 8 (1980) 165-186.
- [4] K.T. Soo, *Regional Science and Urban Economics* 35 (2005) 239-263.
- [5] G.K. Zipf, *Human Behaviour and the Principle of Least Effort*, Addison-Wesley, Cambridge, 1949.
- [6] R.L. Axtell, *Science* 293 (2001) 1818-1820.
- [7] X. Gabaix, *Annual Review of Economics* 1 (2009) 255-294.
- [8] X. Gabaix, Y.M. Ioannides, The evolution of city size distributions. in: V. Henderson, J.F. Thisse, (Eds.), *Handbook of Regional and Urban Economics*, vol. 4. North Holland, Amsterdam 4 (2004) 2341-2378.
- [9] D. Black, J.V. Henderson, *Journal of Economic Geography* 3 (2003) 343-372.
- [10] J. Eeckhout, J., *American Economic Review* 94 (2004) 1429-1451.
- [11] V. Paxsan, S. Floyd, *IEEEACM Transactions on Networking* 3 (1995) 226-244.
- [12] N. Bhattacharya, *Sankhya Ser. B* 25 (1963) 195-196.
- [13] B. M. Hill, *Annals of Statistics* 3 (1975) 1163-1174.
- [14] M.E.J. Newman, *Contemporary Physics* 46 (2005) 323-351
- [15] X. Gabaix, R. Ibragimov, *Journal of Business Economics and Statistics* (forthcoming).
- [16] A. Ades and E. Glaeser, *The Quarterly Journal of Economics* 110 (1995) 195-227.
- [17] E. Rossi-Hansberg, M.L.J. Wright, *Review of Economic Studies* 74 (2007) 597-624.
- [18] C. C. Au, J.V. Henderson, *Journal of Development Economics* 80(2005) 350-388
- [19] G. Anderson, Y. Ge, *Regional Science and Urban Economics* 35 (2005) 756-776.
- [20] L. Gan, D. Li, S. Song, *Economics Letters* 92 (2006) 256-262.

Table 1

Summary statistics of Chinese city size (1,000 persons)

	1999	2000	2001	2002	2003	2004
Mean	483.8	560.2	541.7	573.9	613.0	632.4
Median	182.0	212.1	198.6	223.0	239.5	272.5
Maximum	11,272.2	11,368.2	12,624.1	12,702.2	12,782.3	12,891.3
Minimum	15.0	15.0	14.0	10.0	9.0	16.0
Std. Dev.	926.1	996.0	1,007.6	1,040.4	1,102.0	1,112.9
Urbanization %	25.50	26.17	28.18	29.26	30.83	31.82
Sample size	663	592	664	655	650	654

Sources: *Chinese Urban Statistical Yearbooks* (State Statistical Bureau, 2000-2005)

Table 2

Components of the results of Pareto exponents using the rolling rank regressions

n	1999	2000	2001	2002	2003	2004	Simulation
100	1.8132 ^{***} (0.2564)	1.8290 ^{***} (0.2587)	1.8075 ^{***} (0.2556)	1.8523 ^{***} (0.2620)	1.7807 ^{***} (0.2518)	1.7868 ^{***} (0.2527)	1.0594 0.1498
200	1.5416 ^{***} (0.1542)	1.6336 ^{***} (0.1634)	1.6139 ^{***} (0.1614)	1.6405 ^{***} (0.1640)	1.6108 ^{***} (0.1611)	1.6169 ^{***} (0.1617)	1.0375 0.1037
300	1.2361 ^{**} (0.1009)	1.2574 ^{**} (0.1027)	1.3009 ^{***} (0.1062)	1.3300 ^{***} (0.1086)	1.3114 ^{***} (0.1071)	1.3609 ^{***} (0.1111)	1.0284 0.0840
400	1.0894 (0.0770)	1.0339 (0.0731)	1.0842 (0.0768)	1.0852 (0.0767)	1.0949 (0.0774)	1.1459 [*] (0.0810)	1.0233 0.0724
500	1.0135 (0.0641)	0.9455 (0.0598)	0.9818 (0.0621)	0.9676 (0.0612)	0.9898 (0.0626)	1.0240 (0.0648)	1.0199 0.0645
Full	0.8758 ^{**} (0.0481)	0.8356 ^{***} (0.0486)	0.8291 ^{***} (0.0455)	0.7994 ^{***} (0.0442)	0.8178 ^{***} (0.0454)	0.8546 ^{***} (0.0473)	1.0165 0.0562
Average	1.2809	1.3206	1.2901	1.3094	1.3035	1.3277	1.0358

Notes: Numbers in () are corrected standard errors computed as in Gabaix and Ibragimov [15]. n is the sub-sample size of regression, and the “Full” sample size is 663 for the year of 1999, 592 for 2000, 664 for 2001, 655 for 2002, 650 for 2003, and 654 for 2004. “Average” is the average value for all estimate exponents in each year. “Simulation” is a simulation of an exact rank size rule for 654 cities with the largest city being just the primacy of 2004. *** significant at 1%, ** significant at 5%, * significant at 10%, significantly from one for β . The mean estimate for the “Full” sample of each year is 0.8354.

Table 3

Components of the results of Pareto exponents using the rolling window regressions

truncation point	1999	2000	2001	2002	2003	2004	Simulation
100	1.8132 ^{***} (0.2564)	1.8290 ^{***} (0.2587)	1.8075 ^{***} (0.2556)	1.8523 ^{***} (0.2620)	1.7807 ^{***} (0.2518)	1.7868 ^{***} (0.2527)	1.0594 0.1498
200	0.9939 (0.1406)	1.0873 (0.1538)	1.0415 (0.1473)	1.0812 (0.1529)	1.1010 (0.1557)	1.1156 ^{***} (0.1578)	1.0035 0.1419
300	0.5765 ^{***} (0.0815)	0.4061 ^{***} (0.0574)	0.5481 ^{***} (0.0775)	0.5588 ^{***} (0.0790)	0.5413 ^{***} (0.0766)	0.7011 ^{***} (0.0992)	1.0020 0.1417
400	0.8238 (0.1165)	0.6624 ^{***} (0.0937)	0.6214 ^{***} (0.0879)	0.5198 ^{***} (0.0735)	0.6255 ^{***} (0.0885)	0.5448 [*] (0.0771)	1.0014 0.1416
500	0.7923 [*] (0.1121)	0.6507 ^{***} (0.0920)	0.7149 ^{***} (0.1011)	0.6758 ^{***} (0.0956)	0.6089 ^{***} (0.0861)	0.6412 ^{***} (0.0907)	1.0011 0.1416
End	0.1192 ^{***} (0.0169)	0.1258 ^{***} (0.0178)	0.1170 ^{***} (0.0166)	0.1000 ^{***} (0.0141)	0.0930 ^{***} (0.0132)	0.1153 ^{***} (0.0163)	1.0008 0.1415
Average	0.7698	0.7794	0.7448	0.7330	0.7452	0.7581	1.0028

Notes: Numbers in () are corrected standard errors computed as in Gabaix and Ibragimov [15]. “Truncation point” is the ending point of the fixed-width regression window, and the full sample size is 663 for the year of 1999, 592 for 2000, 664 for 2001, 655 for 2002, 650 for 2003, and 654 for 2004. “End” is the last fixed-width regression window, i.e., the smallest 100 cities in each year. “Average” is the average value for all estimate exponents in each year. “Simulation” is a simulation of an exact rank size rule for 654 cities with the largest city being just the primacy of 2004.

*** significant at 1%, ** significant at 5%, * significant at 10%, significantly from one for β .

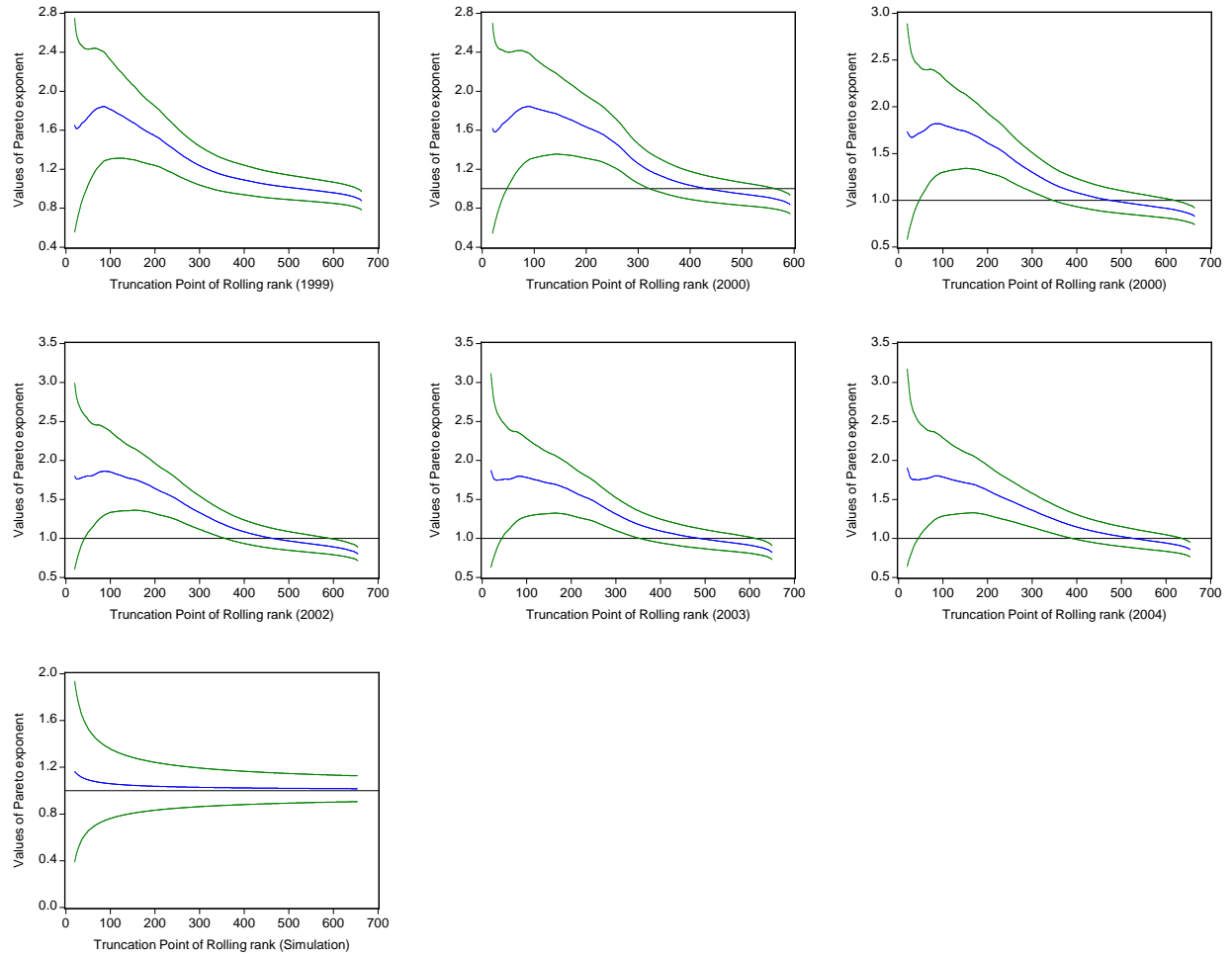


Fig. 1. Values of Pareto Exponent with 95% Confidence Interval using the Rolling Rank Regressions.

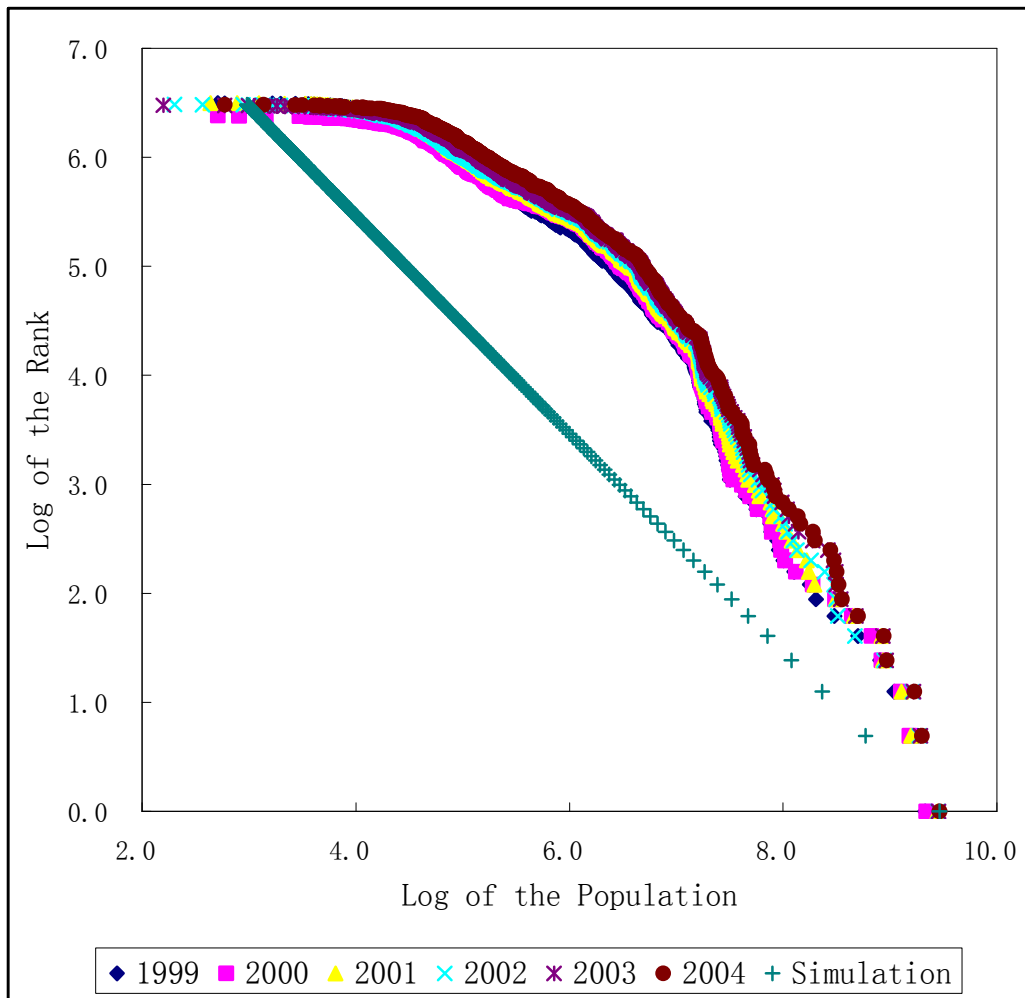


Fig. 2. Log Size versus Log Rank of the Full Sample Size

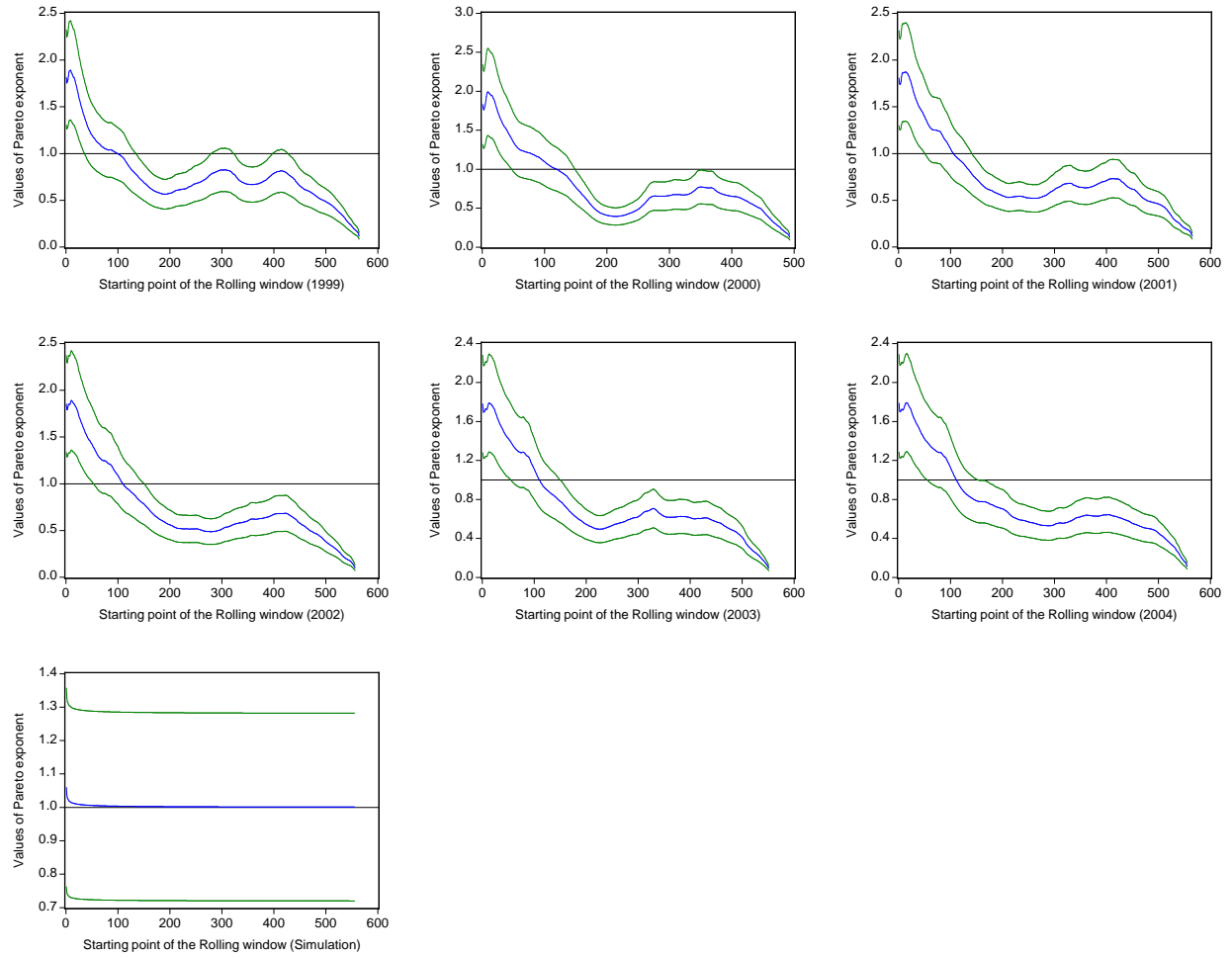


Fig. 3. Values of Pareto Exponent with 95% Confidence Interval using the Rolling Window Regressions.

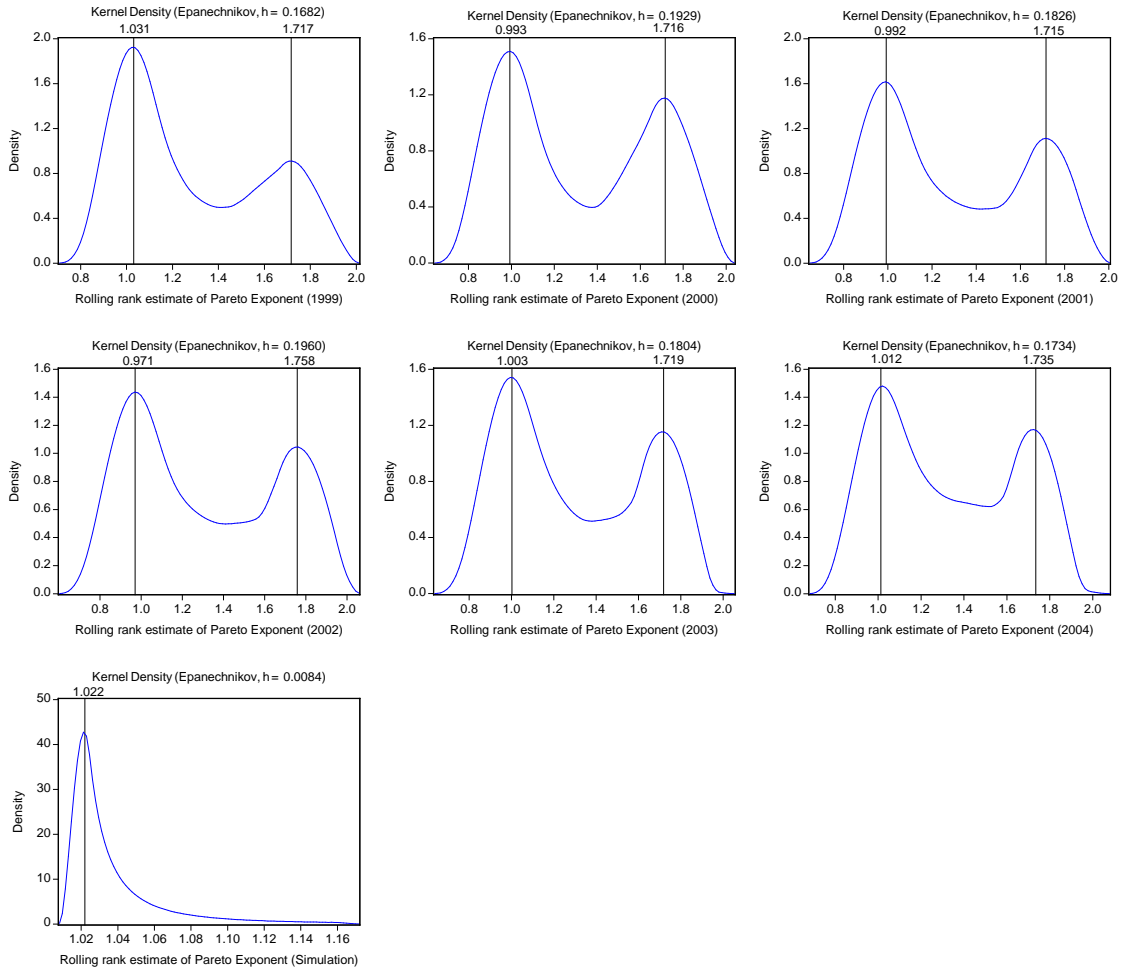


Fig. 4. Kernel Density Function for the Pareto Exponent using the Rolling Rank Regressions.

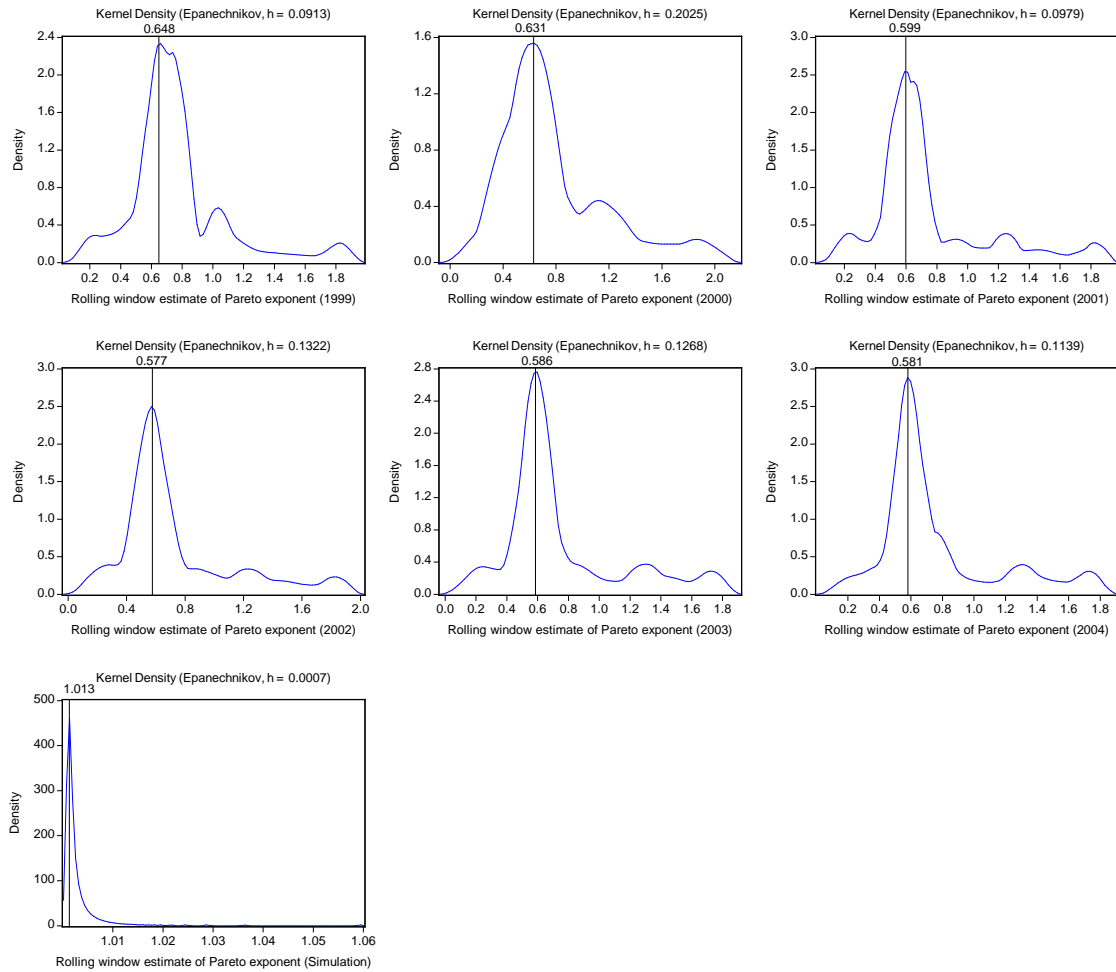


Fig. 5. Kernel Density Function for the Pareto Exponent using the Rolling Window Regressions.

A APPENDIX

Results of the Hill estimator for the Rolling rank method

The Hill estimator is the maximum likelihood estimator under the null of Pareto distribution (Gabaix and Ioannides, 2004). For a sample of n cities with sizes $S_{(1)} \cdots \geq S_{(r)} \geq \cdots \geq S_{(n)}$, Hill estimator is:

$$\beta = \frac{n-1}{\sum_{r=1}^{n-1} (\ln S_{(r)} - \ln S_{(n)})}. \quad (\text{A.1})$$

And the standard error, σ , on β can be computed by the delta method:

$$\sigma = \beta^2 \left(\frac{\sum_{r=1}^{n-1} \left(r(\ln S_{(r)} - \ln S_{(r+1)}) - \frac{1}{\beta} \right)^2}{n-2} \right)^{0.5} (n-1)^{-0.5}. \quad (\text{A.2})$$

In fact, the Hill estimator is almost the same as the MLE estimator discussed in Newman [14]. Comparing with the Hill estimator (equation (7) in Gabaix and Ioannides [8]) with the MLE estimator (equation (B6) in Newman [14]), we can find that they are almost identical, except that the numerator of the Hill estimator is $n-1$ instead of n , which is the numerator of MLE estimator. And, α , the exponent of density function in equation (B6) in Newman [14], is equal to $\zeta + 1$, which is the exponent of (counter) cumulative function in equation (7) in Gabaix and Ioannides [8]. However, as Gabaix [7], Gabaix and Ioannides [8], and Soo [4] point out, both methods (Hill and OLS) have pitfalls. But, the OLS approaches are more robust than the Hill estimator of a tail index under deviations from Pareto laws (see Gabaix and Ibragimov [15]).

Below are the results of the Hill estimator for the Rolling rank method.

We use the Hill estimator for the Rolling rank method with the beginning sub-sample size of 20, an arbitrary start number for each year separately, and get 644 (=663-20+1), 573, 645, 636, 631, and 635 exponent coefficients respectively. The mean value of Pareto exponent for the “full” sample of each year is roughly 0.3203, which is much smaller than the mean OLS estimate of 0.8354 for the “Full” sample of each year. Table A1 below reports the components of the results.

The last column of Table A1 clearly reveals the result for the simulation size series, which

is an exact rank size rule for 654 cities with the largest city being just the primacy of 2004 as described in the end of section 2.2 of the paper. Surprisingly, the values of the Hill estimate of Pareto exponent are significantly different from one for every sub-sample size, although they are so close to 1. This result is contrary to that of OLS.

Figure A1 illustrates the relationship between the Pareto exponent with 95% confidence interval and truncation point from the 20th largest city to the smallest city for each year. The figure suggests that the Pareto exponent is also almost negatively related to the sub-sample size.

We also perform a nonparametric analysis for the distribution of the Hill estimate of Pareto exponent for the Rolling rank method to show a complete description of how the values of the Pareto exponent are distributed, as does Soo (2005). We construct the efficient Epanechnikov kernel function using the optimal window width for the Pareto exponent. Figure A2 demonstrates the kernel function of the Hill estimate of Pareto exponent for the Rolling rank method.

In sum, the results of the Hill estimator are similar to those of OLS. However, the deviations of the Hill estimator are much larger than those of OLS.

Table A1

Components of the results of Hill estimator for the Rolling rank method

n	1999	2000	2001	2002	2003	2004	Simulation
100	1.6439 ^{***} (0.1626)	1.6584 ^{***} (0.1863)	1.6538 ^{***} (0.1677)	1.7100 ^{***} (0.1647)	1.6727 ^{***} (0.1731)	1.6607 ^{***} (0.1850)	1.0230 ^{***} (0.0042)
200	1.2231 ^{***} (0.0902)	1.3331 ^{***} (0.1010)	1.2597 ^{***} (0.0983)	1.2813 ^{***} (0.1070)	1.2544 ^{***} (0.1045)	1.2694 ^{***} (0.0973)	1.0131 ^{***} (0.0022)
300	0.8956 [*] (0.0598)	0.8148 ^{***} (0.0604)	0.9180 (0.0604)	0.9412 (0.0647)	0.9152 (0.0619)	1.0198 (0.0657)	1.0094 ^{***} (0.0015)
400	0.8666 ^{***} (0.0506)	0.7838 ^{***} (0.0476)	0.8121 ^{***} (0.0499)	0.7854 ^{***} (0.0461)	0.8141 ^{***} (0.0470)	0.8343 ^{***} (0.0471)	1.0074 ^{***} (0.0011)
500	0.8393 ^{***} (0.0477)	0.7582 ^{***} (0.0393)	0.7800 ^{***} (0.0437)	0.7453 ^{***} (0.0407)	0.7587 ^{***} (0.0401)	0.7825 ^{***} (0.0410)	1.0061 ^{***} (0.0009)
Full	0.3601 ^{***} (0.0666)	0.3411 ^{***} (0.0560)	0.3410 ^{***} (0.0542)	0.3014 ^{***} (0.0485)	0.2851 ^{***} (0.0547)	0.3332 ^{***} (0.0581)	1.0049 ^{***} (0.0007)
Average	1.0664	1.0983	1.0571	1.0586	1.0569	1.0774	1.0131

Notes: Numbers in () are standard errors computed from equation (A.1). n is the sub-sample size of regression, and the “Full” sample size is 663 for the year of 1999, 592 for 2000, 664 for 2001, 655 for 2002, 650 for 2003, and 654 for 2004. “Average” is the average value for all estimate exponents in each year. “Simulation” is a simulation of an exact rank size rule for 654 cities with the largest city being just the primacy of 2004. *** significant at 1%, ** significant at 5%, * significant at 10%, significantly from one for β . The mean Hill estimate for the “Full” sample of each year is 0.3203, which is much smaller than the mean OLS estimate of 0.8354 for the “Full” sample of each year.

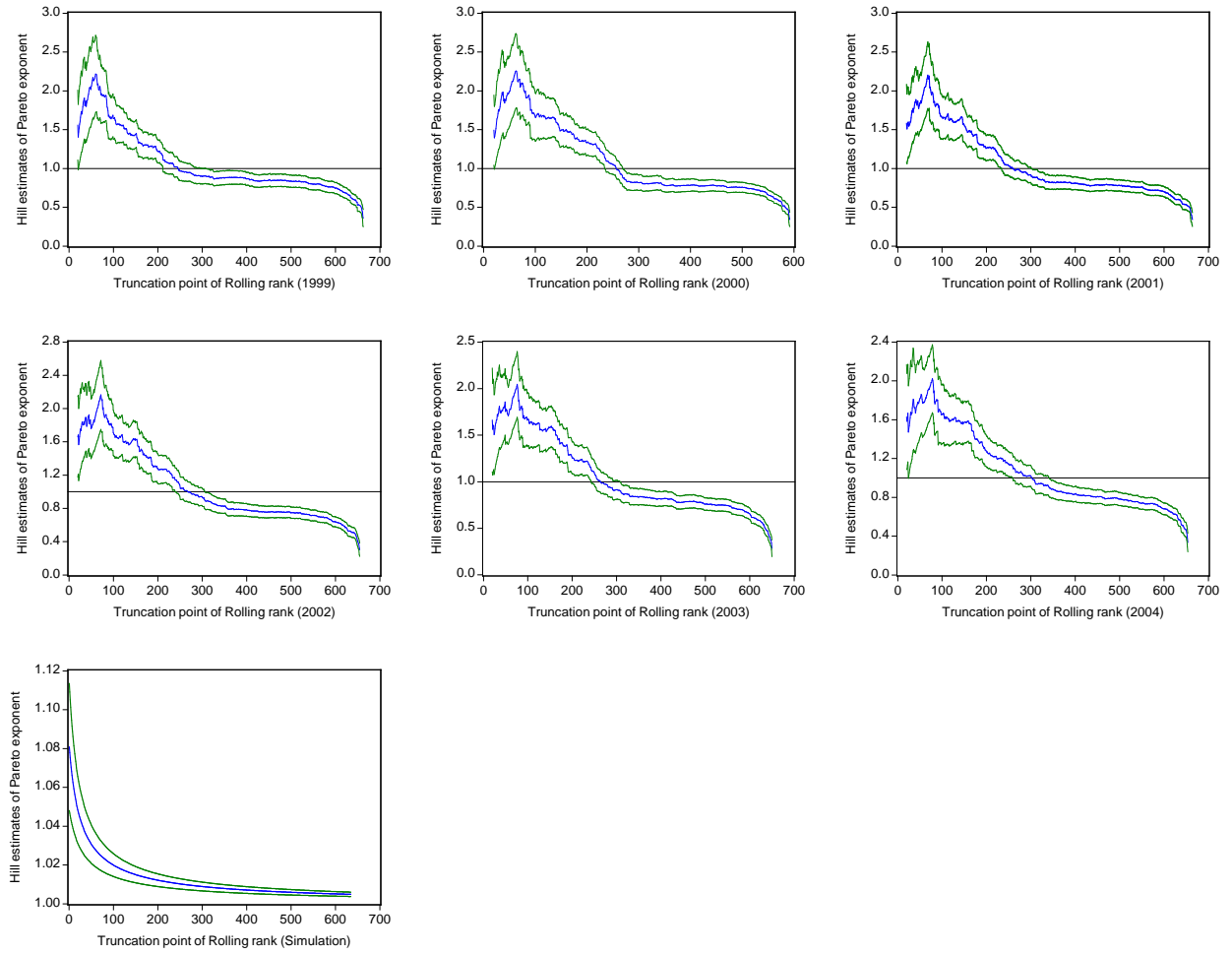


Fig. A1. Values of Hill estimate of Pareto Exponent with 95% Confidence Interval for the Rolling rank method.

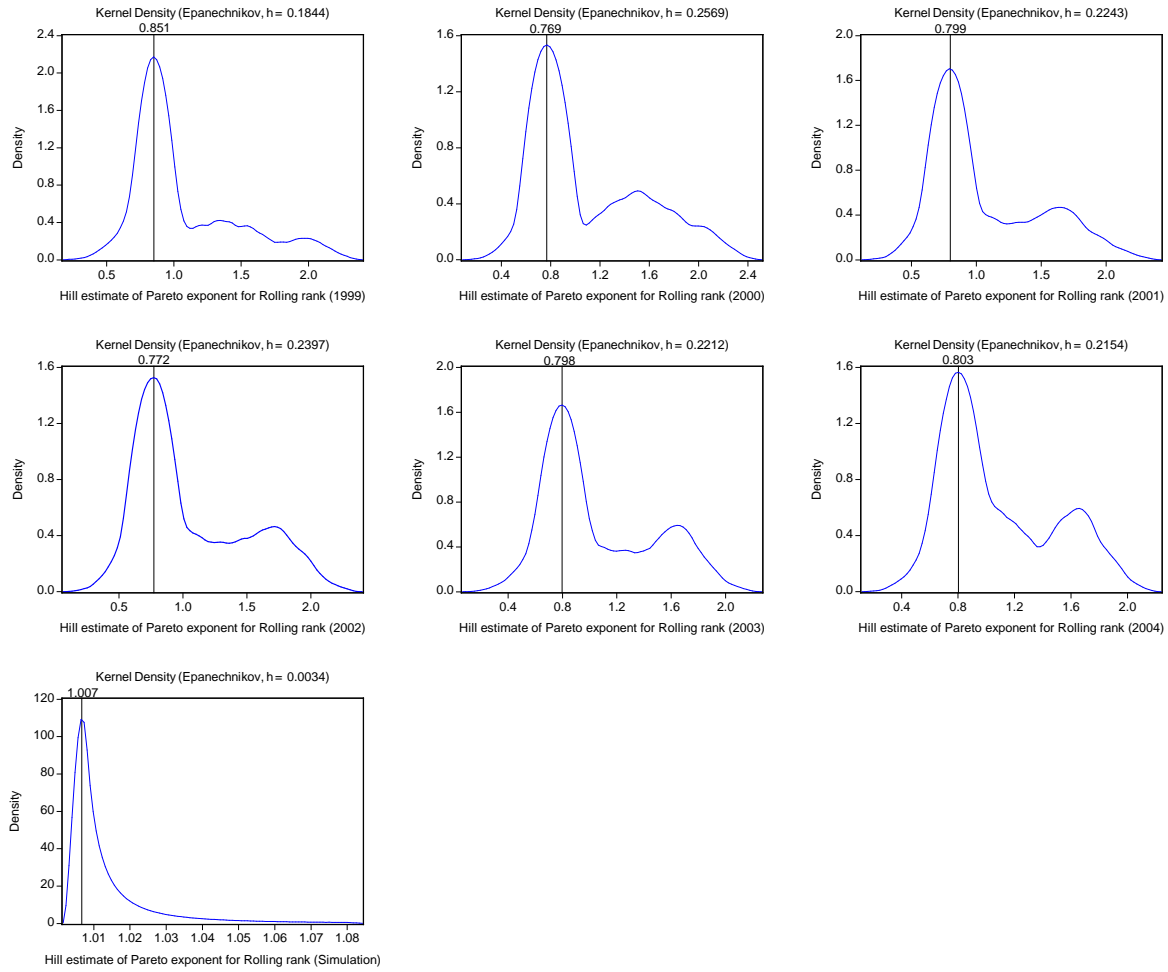


Fig. A2. Kernel Density Function for the Pareto Exponent using the Hill estimator for the Rolling rank method.