Zipf's law, the coherence of the urban system and city size distribution: evidence from Pakistan

Sidra Arshad[a], Shougeng Hu[b,c], Badar Nadeem Ashraf[d,e]

[a]School of Public Administration, China University of Geosciences (CUG), Wuhan 430074, PR China

[b]Department of Land Resources Management, China University of Geosciences (CUG), Wuhan 430074, PR China

[c]Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, PR China

[d]School of Economics and Management, China University of Geosciences Wuhan, Wuhan 430074, China

[e]International School, East China Jiao Tong University, Nanchang 330013, Jiangxi, PR China

Abstract

This paper examines the Zipf's law for the size distribution of Pakistani cities using the five census data from 1951 to 1998. We observe that Zipf's law does not hold in any of the five census years at national level. Next, we consider the city size distribution of four Pakistani provinces and find that Zipf's law is more likely to hold for cities at province-level. We attribute these findings to the coherence property of the urban system. In Pakistan, the urban systems within provinces are more coherent in terms of common language, common culture and common rules as compared to the urban system as a whole at national-level.

Keywords: city size distribution; Zipf's law; Pakistan; coherence

1. Introduction

In this study, we examine the size distribution of Pakistani cities both at the national- and provinces-level using the population census data from 1951 to 1998. Considering the data of all urban localities of the country, first we examine whether the Zipf's law holds at the national-level. Then we examine whether the Zipf's law holds at the regional-level for each of the four provinces of the country. Finally, we discuss the impact of the regional disparity on the city size distribution at the the national-level (the coherence property of the urban system is examined).

In urban studies, Zipf's law has been widely used to describe city size distributions. Zipf's law implies that the city size distribution of a country/region can be described by a Pareto distribution with shape parameter exactly equals to 1. More generally, this law suggests that a country's largest city is approximately twice as large as the second-largest city, three times as large as the third-largest city and so on. Although, the size distribution of cities of a country can be described in alternative ways, such as the measures of primacy [1-4] and alternative functional forms [5-16], however the Zipf's law is more famous among urban scholars. Zipf's law has at least two advantages over the other methods: First, it captures, to some extent, the properties of the whole urban system as compared to the measures of primacy which largely focus more on the largest city. Second, it is easier to interpret the Zipf's law distribution as compared to the other complex functional forms.

Extant literature has examined the empirical validation of Zipf's law for city size distribution and largely reports mixed and controversial evidence. One strand of the literature argues that the size distribution of cities obeys Zipf's law and tries to establish that Zipf's law for city size distribtuion is universal [17-24]. In contrast, another strand of the literature finds that Zipf's law either does not exactly hold [5, 25-30] or the city size distribution may tilt away from Zipf's law over time [31-34] and rejects the hypothesis that Zipf's law is universal (see, for

example, the recent literature survey on Zipf's law and city size distribution by Arshad, Hu and Ashraf [35]). This mixed evidence motivates us to research the topic further. In this context, we provide new empirical evidence using the city data from Pakistan. Pakistan is World's sixth most populous country [36] and has experienced rapid urbanization since its independence in 1947. Despite that several studies have examined Zipf's law for the city size distribution of specific countries, such as the US [17, 19, 22, 28, 37-41], China [23, 25, 42-49], India [23, 43, 44], France [18], Germany [21], Canada [29], Brazil [50-52], Morocco [24], Malaysia [31], Mexico [34] and Turkey [32], it comes to us as a surprise that to date no major effort has been made to study the city size distribution of Pakistan.

Besides this mixed evidence, the controversy arises with the results of both strands of the literature. The findings of the studies which support Zipf's law are questioned on the basis of sample selection, while the results of the studies which reject Zipf's law are criticized on the basis of the properties of the urban system to which cities belong.

Almost all of the studies which favor Zipf's law examine the size distribution of upper tail cities only, largely excluding the data of small cities below a truncation point. In this context, one major drawback is that the choice of the truncation point lies with the researcher and causes arbitrariness in the sample selection. A number of recent studies have observed that the estimated value of Pareto exponent is sensitive to the sample size [5, 31, 37, 53, 54]. In a very recent study, Fazio and Modica [37] show that the estimated value of Pareto exponent decreases as the smaller size cities are included into the sample. One implication is that a certain sample of large cities can be selected to support the hypothesis of Zipf's law for the city size distribution when the actual city size distribution is not Zipf's law. In this study, we eliminate the sample selection concern by using the data of all urban localities of Pakistan for empirical analysis.

On the other hand, the studies which reject the Zipf's law are criticized on the basis of the 'coherence' property of the urban system. Cristelli, Batty and Pietronero [55] argue that mixed and controversial empirical evidence is due to the very definition of the objects comprising the system to which Zipf's law is applied. They suggest that a system of cities does not obey true power law behavior because it is either incomplete or inconsistent with the conditions under which one might expect power laws to emerge. Power laws can only be applied to a group of cities which are integrated institutionally (i.e., common rules, common culture, common language, etc.) and economically and have co-evolved over time. The group of cities, which historically has observed an integrated evolution, converges to an organic economic unit. Consequently, the size distribution of cities becomes internally consistent for the group as a whole and obeys statistical properties of power laws. In this perspective, some recent studies focus on economic factors and argue that a dynamic interplay between economic activity and growth process of cities results Zipf's law [38, 56-62]. In this study, we examine the city size distribution of Pakistan with respect to the coherence property of the urban system. Pakistan came into being in 1947 after independence from British colonial rule. It has an economic integration at national level in addition to a considerable internal diversity with four provinces which have entirely different resource endowment, cultures, and languages. With these characteristics, Pakistan offers an ideal laboratory to examine the impact of the coherence of the urban system on the size distribution of cities.

The rest of the paper proceeds as follow: Section 2 introduces the Zipf's law and methodology used in empirical analysis. Section 3 introduces the dataset used in the study and the urbanization landscape of Pakistan. Section 4 reports empirical results. And, the final section concludes the findings of the paper.

2. Introduction to Zipf's Law and methodology

Auerbach [63], Singer [64] and Zipf [65] have established that the city size distribution can be represented as a Pareto distribution.

$$R = A.P^{-\alpha} \qquad Eq.(1)$$

Or

$$Log\ (R) = Log\ (A) - \ \alpha.Log\ (P) \qquad Eq.(2)$$

Here, *P* represents the population of a city. *R* is the rank of a city when cites are ranked from 1 to n by the population size. A and α are constants. α is also referred as Pareto exponent.

City size distribution is said to follow Zipf's law if the estimated value of α in *Eq.* (2) is statistically equal to 1. In that case, the plot of the log of rank versus the log of city size shows a scatter diagram with a regression line having slope equal to -1.

A general approach in city size distribution literature is to use ordinary least squares (OLS) regression to estimate the α in *Eq.* (2). However, Gabaix and Ioannides [66] use Monte Carlo simulations and show that OLS estimates of α in *Eq.* (2) are downward biased in small samples due to serial correlation caused by the ranking procedure used to generate R (i.e., dependent variable). Similarly, another problem with the OLS estimation of Eq. (2) is that the standard errors would be underestimated and, as a result, Zipf's law can be rejected too often based on the *t*-test [67, 68]. To correct for these biases linked with the estimation of Eq. (2), Gabaix and Ibragimov [69] suggest to use 'Rank-1/2' as dependent variable and to estimate the following variation of *Eq.* (2).

$$Log\ \left(R - \frac{1}{2}\right) = Log\ (A) - \ \alpha.Log\ (P) \qquad Eq.(3)$$

To estimate the results for this study, we follow the guidelines of Gabaix and Ibragimov [69] and use the Rank-1/2 as dependent variable and estimate the log(rank-1/2) versus log(population) regression as stated in the form of Eq. (3) to estimate the values of α. Then we use an ordinary *t*-test to examine the null hypothesis that the estimated value of α from Eq. (3) equals 1, i.e., α=1. To apply t-test, usually the normality assumption is assumed for the standard errors estimated from Eq. (3), but this is restrictive in some cases. A variable which has been transformed by taking the natural log yields a distribution that is closer to normality. In addition, normality has little impact on results for large samples, i.e., N>20 [70].

3. Data collection and the urbanization landscape of Pakistan

3.1 Data collection

The definition of a city has captured special attention in Zipf's law and city size distribution research. In this context, recent literature has shown that the applicability of Zipf's law is sensitive to the definition of cities [5, 22, 26, 71-78]. Broadly, three types of city definitions have been used in literature: administratively defined cities [5], functionally defined cities [22, 76-78] and natural cities [72-74]. In this study, we rely on the administrative definition of cities which has been used in population censuses where a settlement is considered urban area if it has a local administrative government.

The data for urban populations and cities of Pakistan is collected from the website of Pakistan Bureau of Statistics (PBS) available at the link http://www.pbs.gov.pk/population-tables. PBS uses four level administrative units to measure urban localities: metropolitan corporations, municipal corporations, municipal committees (MC), town committees (TC) and cantonments. Metropolitan cities are the largest cities including Karachi, Lahore and the Capital

Territory Islamabad. Municipal corporations are also for relatively big cities, which are not mega cities like Karachi or Lahore. Municipal committees are for medium sized cities, while the small cities have town committees. Some cantonment areas are largely part of any of the metropolitan, large or small city, while some cantonment areas are considered as standalone cities. When a cantonment area is a part of any other city, then its population is added to the population of that city for the calculation of total population of the urban area. Table (1) reports the total number of cities at national-level and for each of the four provinces in five census years. After the 1998 census, a recent census has been carried out in Pakistan in April 2017 however the data of this latest census is not available yet.

Table 1: Number of cities of whole Pakistan and each of four provinces

Census year	Pakistan	Balochistan	Khyber Pakhtunkhwa	Punjab	Sindh
1951	205	16	21	139	29
1961	276	21	31	160	64
1972	332	23	31	183	94
1981	383	27	34	203	118
1998	474	41	44	237	151

Note: Data is collected from the website of Pakistan Bureau of Statistics (PBS). The column 'Pakistan' includes Islamabad (the Capital of Pakistan) in the years 1972, 1981 and 1998.

3.2 Urbanization landscape of Pakistan

Pakistan is located in South Asia with a total area of 796096 square kilometers. Map in Figure (1) shows the location and administrative units of Pakistan. The history of current Pakistan dates back to 1947 when British rule ended, and Indian sub-continent was partitioned into two countries: Pakistan and India. At the time of partition, Pakistan was consisted of two parts: East Pakistan (today's Bangladesh) and West Pakistan (today's Pakistan). Later in 1971, East Pakistan separated from the West Pakistan (from hereafter Pakistan) and became the independent country under the name of Bangladesh.

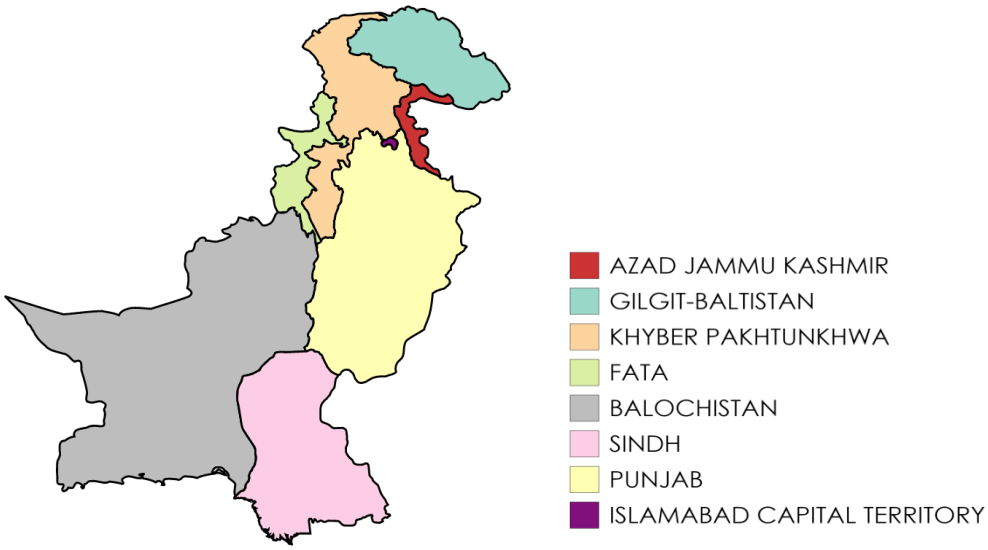


Figure 1 Sub-regions (including four main provinces) of Pakistan

Since 1947, five population censuses have been carried out in Pakistan in the years 1951, 1961, 1972, 1981 and 1998. The country has experienced rapid urbanization since its inception. While the total population quadrupled from 33.7 million in 1951 to 132.3 million in 1998, the urban population had increased 7.2 times from 5.98 million in 1951 to 43 million in 1998 as shown in Table 2. This suggests that the percentage of total population living in cities increased at a quite rapid pace as compared to the overall growth in country's population. As shown in Table 2, the percentage of urban population grew from 17.74 percent in 1951 to 32.51 percent in 1998. The number of cities, as well as, the average size of cities has increased from 1951 to 1998.

Table 2: Total, rural and urban populations of Pakistan

Census year	Total population	Rural population	Urban population	Urban population (percentage)	Intercensal urban population growth (percentage)
1951	33,740,167	27,754,670	5,985,497	17.74	-
1961	42,880,378	33,225,806	9,654,572	22.51	61
1972	65,309,340	48,715,689	16,593,651	25.41	72
1981	84,253,644	60,412,173	23,841,471	28.30	43

1998	132,352,279	89,315,875	43,036,404	32.51	80

Note: Data is collected from the website of Pakistan Bureau of Statistics (PBS).

At regional level, Pakistan has four main provinces: Balochistan, Khyber Pakhtunkhwa, Punjab and Sindh. This provincial distribution is more natural and historic as each province largely has evolved with different regional language and culture. The movement within provinces is easier for population than across the provinces due to the cross-provincial differences. There are some exceptions, such as, in the case of biggest city of Karachi or in the case of capital city of Islamabad where migration from all around the Pakistan occurs. Rural-urban distribution of population for four provinces is given in Table 3.

Table 3: Total, rural and urban populations of four provinces of Pakistan

Census year	Total population	Rural population	Urban population	Urban population (Percentage)
		Balochistan		
1951	1,167,167	1,022,618	144,549	12.39
1961	1,353,484	1,125,016	228,468	16.88
1972	2,428,678	2,029,094	399,584	16.45
1981	4,332,376	3,655,604	676,772	15.62
1998	6,565,885	4,997,105	1,568,780	23.89
		Khyber Pakhtunkhwa		
1951	4,556,545	4,051,800	504,745	11.08
1961	5,730,991	4,972,475	758,516	13.23
1972	8,388,551	7,192,896	1,195,655	14.25
1981	11,061,328	9,395,675	1,665,653	15.06
1998	17,743,645	14,749,561	2,994,084	16.87
		Punjab		
1951	20,540,762	16,972,686	3,568,076	17.37
1961	25,463,974	19,988,052	5,475,922	21.50
1972	37,607,423	28,424,728	9,182,695	24.41
1981	47,292,441	34,240,795	13,051,646	27.60
1998	73,621,290	50,602,265	23,019,025	31.27
		Sindh		
1951	6,047,748	4,279,621	1,768,127	29.24
1961	8,367,065	5,200,047	3,167,018	37.85

1972	14,155,909	8,430,133	5,725,776	40.45
1981	19,028,666	10,785,630	8,243,036	43.32
1998	30,439,893	15,600,031	14,839,862	48.75

Note: Data is collected from the website of Pakistan Bureau of Statistics (http://www.pbscensus.gov.pk/content/population-and-housing-indicators).

There are, at least, two reasons of rapid urbanization in Pakistan: internal migration and international/external migration.

According to the latest census of 1998, 10.8 million (8 percent) are migrants out of total population of 132 million as shown in Table 4. A person is counted as migrant if he is currently residing in a tehsil or district other than the one in which he born. This definition includes only internal migrants. Of all migrants, 64 percent have moved to cities. There are multiple reasons of internal migration to cities. According to the 1998 census, 43 percent of all lifetime migrants said that they had moved with the household head, 17 percent because of marriage, 12 percent for employment and 9 percent for business. People have also moved to cities due to poor law and order situation and scarce educational opportunities especially for the girls in rural areas [79, 80].

Table 4: Migrant population according to the 1998 census year

Present residence	Total migrant population (1998 census year)	Total migrant population (Percentage)
All Areas	10,829,264	100%
Khyber Pakhtunkhwa	647,725	5.8%
Punjab	6,701,256	61.9%
Sindh	2,833,227	26.3%
Balochistan	249,615	2.3%
Islamabad	397,731	3.7%

Note: Data is collected from the website of Pakistan Bureau of Statistics (http://www.pbscensus.gov.pk/content/population-and-housing-indicators).

In addition to internal migration from rural areas to urban centers, the urban landscape of Pakistan has observed three major external shocks: the migration from India in 1947 at the time of partition, the migration due to three wars in 1948, 1965 and 1971 between India and Pakistan, and the migration of refugees from Afghanistan due to Soviet invasion in 1979.

At the time of partition in 1947, 6.5 million Muslims migrated from India to Pakistan while 4.7 million Hindus and Sikhs left from today's Pakistan to India [81]. This widespread migration changed the landscape of whole country including all provinces. Urban population increased between 90 to 192 percent during the inter-census period of 1941 (the last pre-partition census) to 1951 for many cities in Punjab due to refugees settlement. These cities are largely located in the districts of Bahawalnagar, Rahim Yar Khan, Faisalabad and Toba Tek Singh. Comparatively, these cities grew much less during the inter-census period of 1931-1941. On the other hand, several cities located in the districts of Dera Ghazi Khan, Layyah and Rajanpur registered a negative growth due to the departure of Hindus and Sikhs. Similarly, the urban population of Sindh increased from 12 percent in 1941 to 29 percent in 1951. Again the trend was not homogeneous; the sizes of several small cities declined due to the departure of Hindus, while the sizes of two major cities, Hyderabad and Karachi, increased almost 150 percent due to incoming refugees' settlement. For Khyber Pakhtunkhwa, the percentage of urban population decreased from 18 percent in 1941 to 11 percent in 1951 due to the departure of Hindus and the tendency of incoming refugees to not settle in Khyber Pakhtunkhwa but in Punjab and Sindh. The situation in Balochistan province was also similar to Khyber Pakhtunkhwa.

The migration to Pakistan has also occurred due to three major wars between India and Pakistan in 1948, 1965 and 1971. The Kashmir war of 1948 caused Kashmiries to settle in other urban areas of the Pakistan especially the large scale migration to Karachi. Similarly, during the

1965 and 1971 wars, several Hindu families fled to India while round 3,500 Muslim families moved from Indian Thar to Pakistani Thar and were provided with 42,000 acres of land. Much of these families settled in the small urban centers of Thar Desert.

Another major shock hit the urban landscape of the country during the Soviet invasion of Afghanistan. Nearly 3.7 million Afghan refugees migrated to Pakistan in 1979 and later due to Soviet invasion and largely settled in big cities of Khyber Pakhtunkhwa and Balochistan. As a result, these cities registered abnormally high growth rates during this period. For example, according to the census reports, the average annual urban population growth rate of Peshawar, the capital of Khyber Pakhtunkhwa, rose from 1.9 percent a year during the inter-census period of 1961-1972 to 9.2 percent during the inter-census period of 1972-1981, and again fell to 3.3 percent a year during the inter-census period of 1981-1998. Similarly, the average annual urban population growth rate of Quetta, the capital of Balochistan, rose from 3.44 percent a year during the inter-census period of 1961-1972 to 7.2 percent during the inter-census period of 1972-1981, and again fell to 4.04 percent a year during the inter-census period of 1981-1998.

These migration patterns might have important impact on city size distribution of the country. For example, Simon [82] characterized the relationship between migration and Zipf's law for cities. He argued the size distribution of cities of a region would only obey Zipf's law if the migration to or from cities is proportional to city size; that is, net addition or loss of population of individual cities within the region was proportional to city size.

4. Results

This section reports empirical results. For empirical analysis, first we estimate Pareto exponent at national level and then for each of the four provinces of the country. Finally, we explain the deviations from Zipf's law in terms of the coherence property of the urban system.

Mainly, we use Equation (3) to estimate the Pareto exponent 'α'. The estimated values of α would show the deviations from Zipf's law. A value statistically equal to 1 would confirm that the city size distribution obeys a strict form of Zipf's law. A value statistically less than 1 indicates the city size distribution is more uneven where one or a few cities dominate the whole urban structure. This is also referred as primatial or macrocephalous distribution. On the contrary, a value statistically higher than 1 indicates a more even city size distribution where the size difference between larger and smaller cities is little. For empirical results, first we estimate α at national-level considering the data of all cities to examine whether Zipf's law applies at country-level. Then we estimate α for the cities of each of four provinces to examine the validation of Zipf's law at province-level. Though we rely on the estimates of α from Eq. (3) for main conclusions, however we draw a scatter plot (from hereafter rank vs. size plot or Zipf's curve) with the log of ranks-1/2 on the y-axis and the log of population of cities on the x-axis for a visual representation. The points on the Zipf's curve should conform to a line with a slope of -1 to confirm the strict form of Zipf's law.

4.1 The city size distribution of entire country

First, the Zipf's curve is drawn for all five census years to show the overall trend in the distribution as shown in Figure (2). The Zipf's curve for the census year 1951 shows discontinuities and clear deviations from the linear fit line. Two largest cities, Karachi and Lahore, at the lower end of the distribution were too large in 1951 far away from the fitted line. Similarly, there are other groups with discontinuities and deviations especially the group of next seven cities and then other small towns. These deviations can be attributed to the partition of the subcontinent in 1947. Before that, all cities were part of greater British India, but during the partition, a large proportion of Muslim population migrated from India to Pakistan, while the

Hindu population migrated from Pakistan to India. Partition on the one hand and the two-way migration, on the other hand, changed the urban landscape of the whole region because most of the migrated population settled in the major cities. Zipf's curve of the year 1998 shows that urban structure has evolved towards more linear distribution. However, it is still not completely linear with slight deviations from the linear fitted line. Figure (3) shows the Zipf's plots for each census year with fitted lines and 95% confidence intervals. Actual data points in all five census years deviate from fitted linear line. These plots to some extent suggest that the strict form of Zipf's law seems to be not confirmed for the size distribution of Pakistani cities.

Next, we use Eq. (3) to estimate Pareto exponent for each of the five census years. As shown in Table 5, the estimated values of Pareto exponent in all census years are lower than 1. These results show that size distribution is more uneven than the predicted by the Zipf's law. These results are consistent with the graphical representation which has shown clear deviations from the linear line. The estimated value of Pareto exponent increased from its lowest value (0.91) in 1951 to its highest value (0.94) in 1981 and then again decreased to 0.91 in 1998. The hypothesis that Pareto exponent equals to 1 is rejected for all census years. As shown in Figure (4), the 95% confidence intervals of the point estimates of Pareto exponent in all five census years are lower than 1. Together, these results show that Zipf's law does not hold for Pakistan.

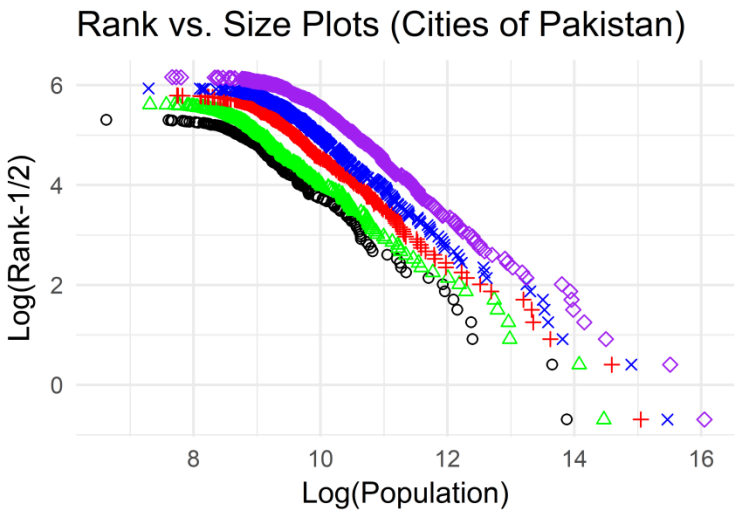


Figure 2: Log (rank-1/2) vs. log (Population) plots for all cities of Pakistan (Black to purple colored plots show Zipf's curves for the census year 1951, 1961, 1972, 1981 and 1998, respectively).

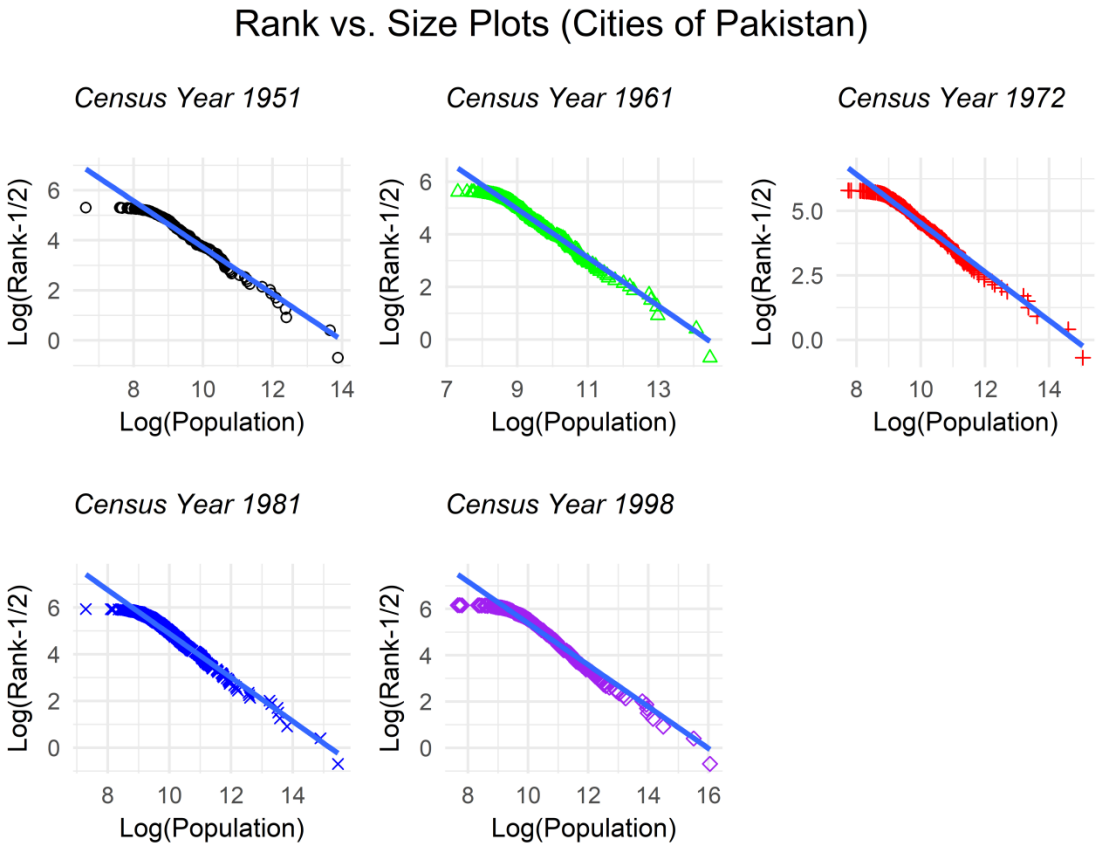


Figure 3 Log (rank-1/2) vs. log (Population) plots for five census years at national-level with fitted lines

Table 5 Log(Rank-1/2) vs. log(Population) regression results at national-level using the data of all cities of Pakistan

Census Year	Number of cities	*Log (A)*	α	*t-test $\alpha=1$	P-value of t-test
1951	205	12.988	0.927	Rejected	0.000
1961	276	13.236	0.919	Rejected	0.000
1972	332	13.968	0.944	Rejected	0.000
1981	383	14.241	0.936	Rejected	0.000
1998	474	14.382	0.900	Rejected	0.000

*95% Confidence interval is used to test hypothesis

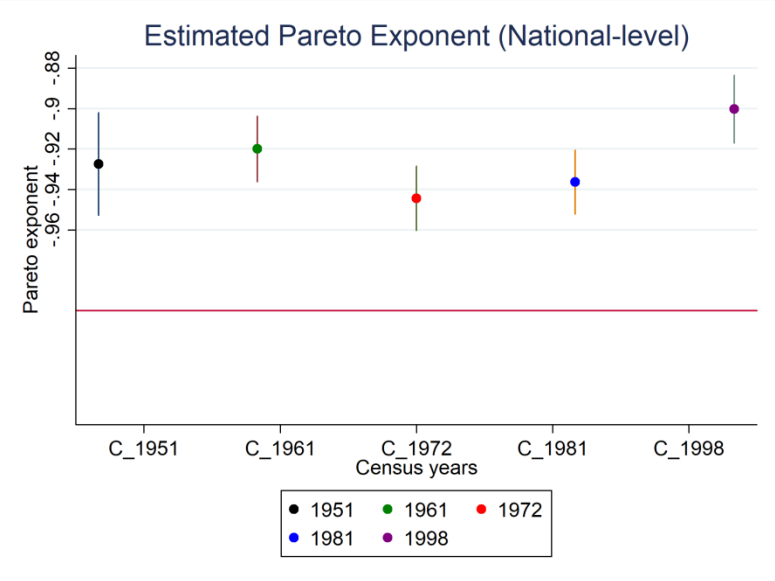


Figure 4 Estimated Pareto exponent at national-level with 95% confidence intervals

4.2 The city size distribution of provinces

Next, we examine the city size distribution at province level. Specifically, we investigate whether the Zipf's law holds in each of the four provinces (Punjab, Sindh, Khyber Pakhtunkhwa and Balochistan) of Pakistan. We use Eq. (3) to estimate the α for the cities of each of the four provinces in each census year. We observe that the estimated values of α for provinces are more likely to be statistically equal to 1; that is, the Zipf's law is more likely to hold in provinces.

Figure (5) shows the Zipf's curves, Table (6) reports the results of log(rank-1/2) log(population) regressions and Figure (6) shows the statistical significances of the estimated Pareto exponents for each of the five census years for the cities of Punjab province. Figure (7) shows that Zipf's curves are more linear as compared to the curves at national-level in Figure (5) and show a better fit to the linear fitted line. Similarly, the estimated values of Pareto exponent for the cities of Punjab province are statistically equal to one in the census years 1961, 1972 and 1981 as shown in Table (6). The estimated values of Pareto exponent have decreased from 1.028 in 1951 to 0.963 in 1998 representing that city size distribution of the province is becoming more uneven over time. More even distribution in 1951 seems an outcome of widespread migration in 1947 when incoming refugees largely settled in small cities of the Punjab province. However, the government policy to consistently focus on large cities such as Lahore, Rawalpindi and Faisalabad has increased their size over time resulting in lower values of Pareto exponent.

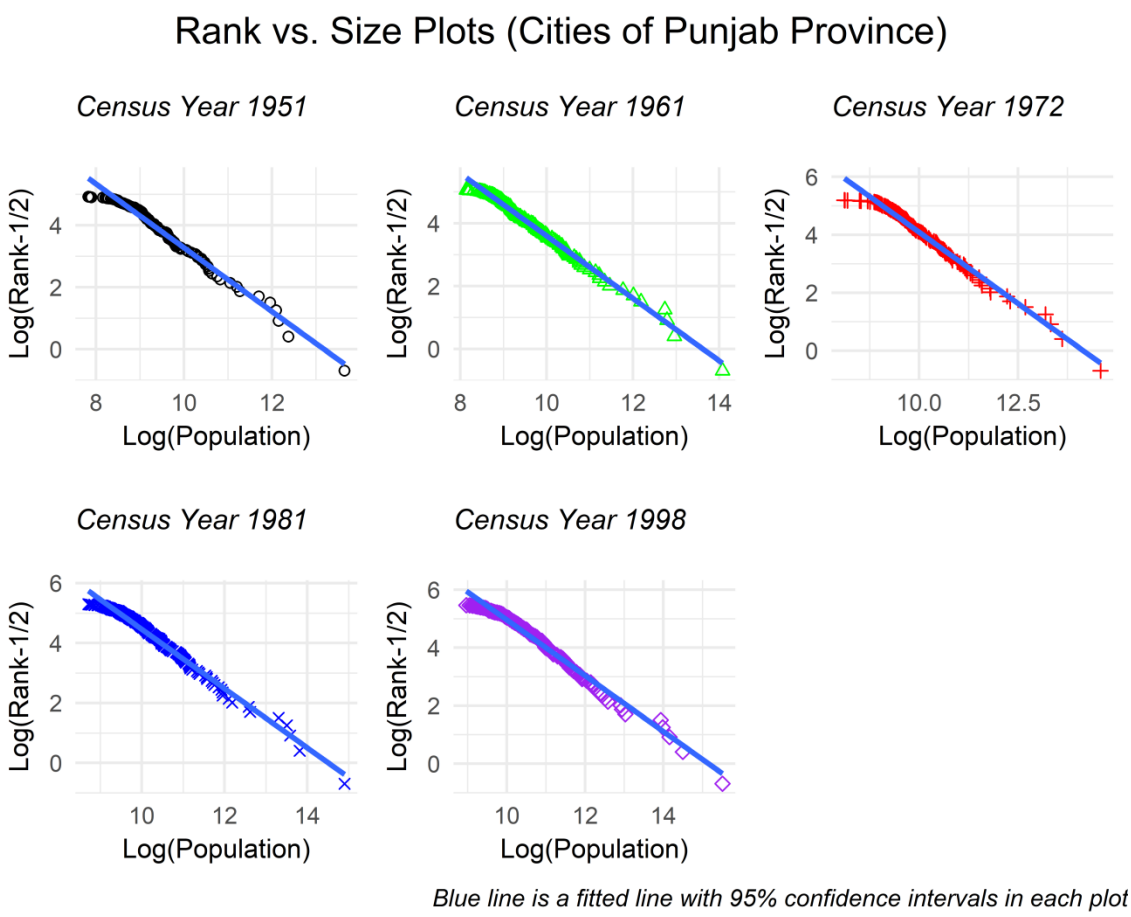


Figure 5 Log (rank-1/2) vs. log (Population) plots for the cities of Punjab Province

Table 6 Log(Rank-1/2) vs. log(Population) regression results for the cities of Punjab province

Census Year	Number of cities	*Log (A)*	α	*t-test $\alpha=1$	P-value of t-test
1951	139	13.557	1.028	Rejected	0.021
1961	160	13.536	0.994	Do not reject	0.414
1972	183	13.932	0.984	Do not reject	0.132
1981	203	14.358	0.990	Do not reject	0.218
1998	237	14.596	0.963	Rejected	0.000

*95% Confidence interval is used to test hypothesis

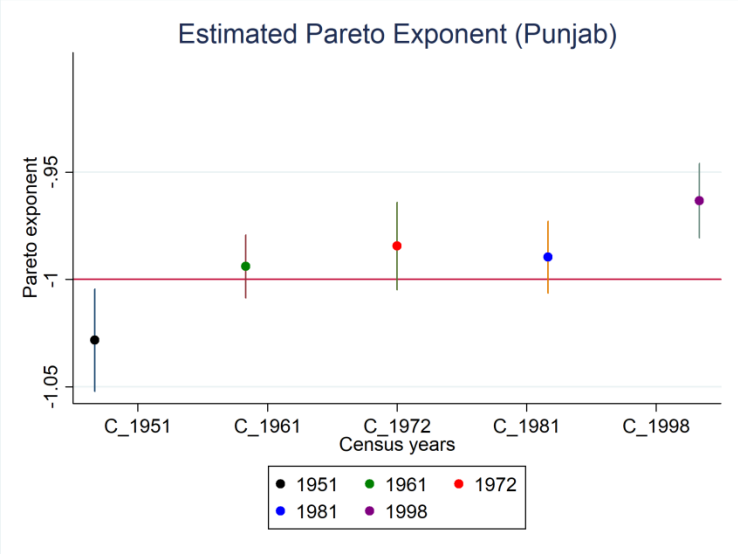


Figure 6 Estimated Pareto exponent (with negative sign as given in Eq. 3) for the cities of Punjab Province with 95% confidence intervals

Figure (7) shows Zipf's curves, Table (7) reports the results of log(rank-1/2) log(population) regressions and Figure (8) shows the statistical significances of the estimated values of α for each of the five census years for the cities of Sindh province. The estimated values of α for the cities of Sindh province are not statistically equal to 1 in any of the five census years as shown in Figure (8). One expected reason of this much uneven city size distribution in Sindh province is the mega city of Karachi. Karachi had remained capital of Pakistan from 1947 to 1960 and is the only port and mega city. Due to enormous employment opportunities, it has attracted migrants not only from all around the country but also a large number of international migrants at the time of partition in 1947 and later on Afghan refugees during Afghan-Soviet War. As shown in Zipf's curves in Figure (7), the distribution is largely linear except that the biggest city, Karachi, at lower end is quite big. In our opinion, Karachi is an outlier in the city system of Sindh province and might bias the results. To account for this concern, we repeated the regression results by dropping the data of Karachi city from the sample.

As shown in Figure (9), the city size distribution of the province is more likely to be Zipfian as Pareto exponent is statistically equal to 1 in some census years. Though the estimated values of Pareto exponent are very near to 1 after dropping the data of Karachi, however the values are not stable due to the absence of main urban centre from the city size distribution of the province.

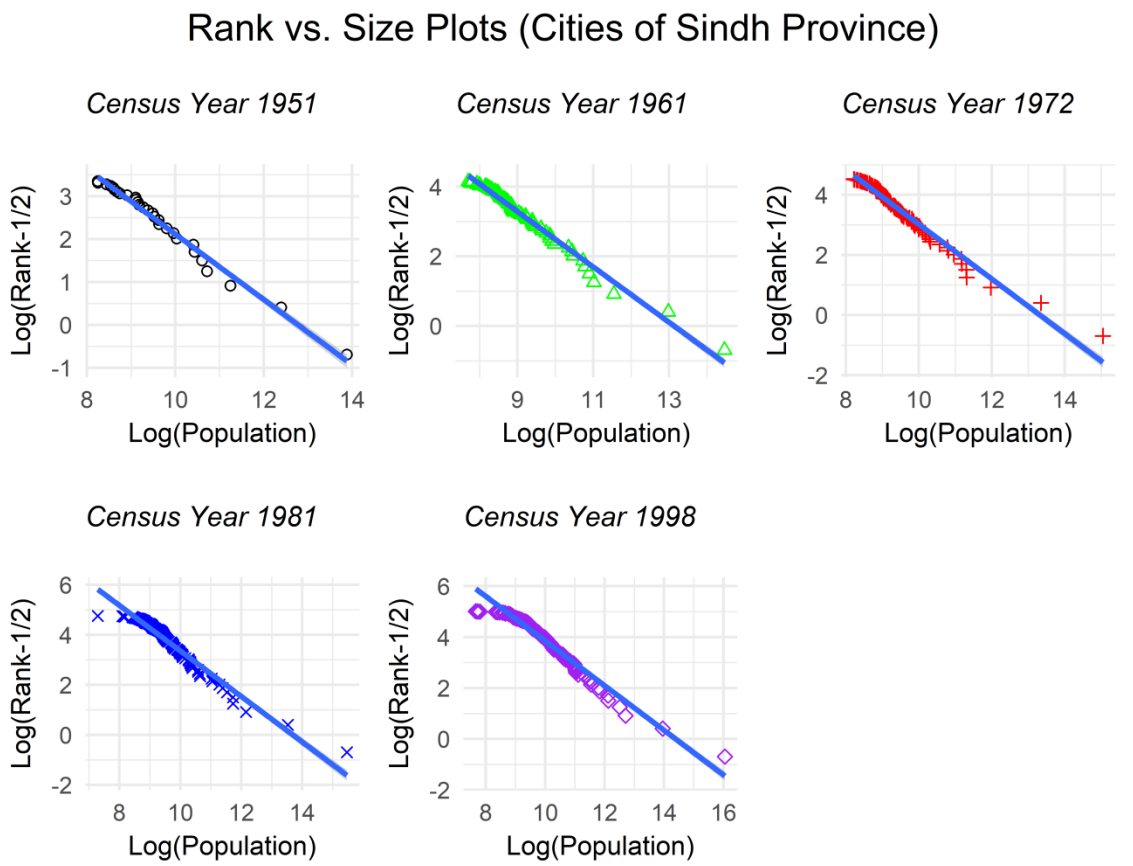


Figure 7 Log (rank-1/2) vs. log (Population) plots for the cities of Sindh Province

Table 7 Log(Rank-1/2) vs. log(Population) regression results for the cities of Sindh province

	Include Karachi					*Exclude Karachi*				
Census Year	Number of cities	*Log (A)*	*α*	*t-test *α=1*	P-value of t-test	Number of cities	*Log (A)*	*α*	*t-test *α=1*	P-value of t-test
1951	29	9.738	0.762	Rejected	0.000	28	11.632	0.977	Do not reject	0.456
1961	64	10.438	0.795	Rejected	0.000	63	11.718	0.945	Rejected	0.001
1972	94	12.118	0.910	Rejected	0.000	93	13.770	1.093	Rejected	0.000
1981	118	12.450	0.909	Rejected	0.000	117	13.837	1.061	Rejected	0.004

1998	151	12.611	0.876	Rejected	0.000	150	13.593	0.981	Do not reject	0.310

*95% Confidence interval is used to test hypothesis

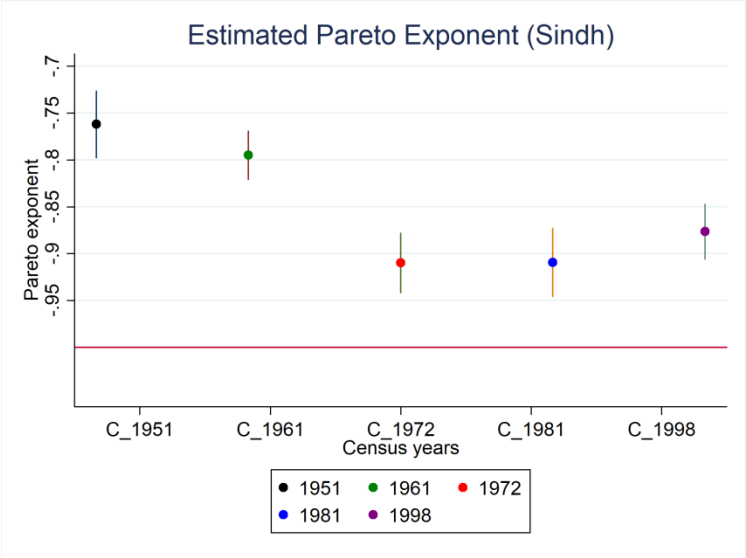


Figure 8 Estimated Pareto exponent (with negative sign as given in Eq. 3) for the cities of Sindh Province with 95% confidence intervals

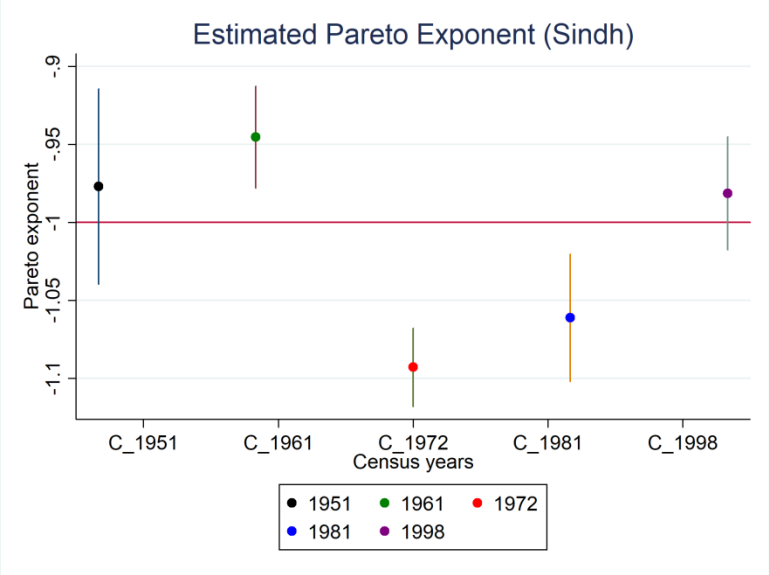


Figure 9 Estimated Pareto exponent (with negative sign as given in Eq. 3) for the cities of Sindh Province (excluding Karachi) with 95% confidence intervals

Figure (10) shows Zipf's curves, Table (8) reports the results of log(rank-1/2) log(population) regressions and Figure (11) shows the statistical significances of the estimated Pareto exponents for each of the five census years for the cities of Khyber Pakhtunkhwa province. As shown in Figure (11), the estimated values of Pareto exponent for the cities of Khyber Pakhtunkhwa province are statistically equal to 1 in all census years. These results show that the city size distribution of the Khyber Pakhtunkhwa province follows Zipf's law. One reason is that though the urban population decreased due to the departure of Hindus from all cities of the province during the partition in 1947, however incoming refugees largely did not settle in the province which had not caused any major distortion in city size distribution.

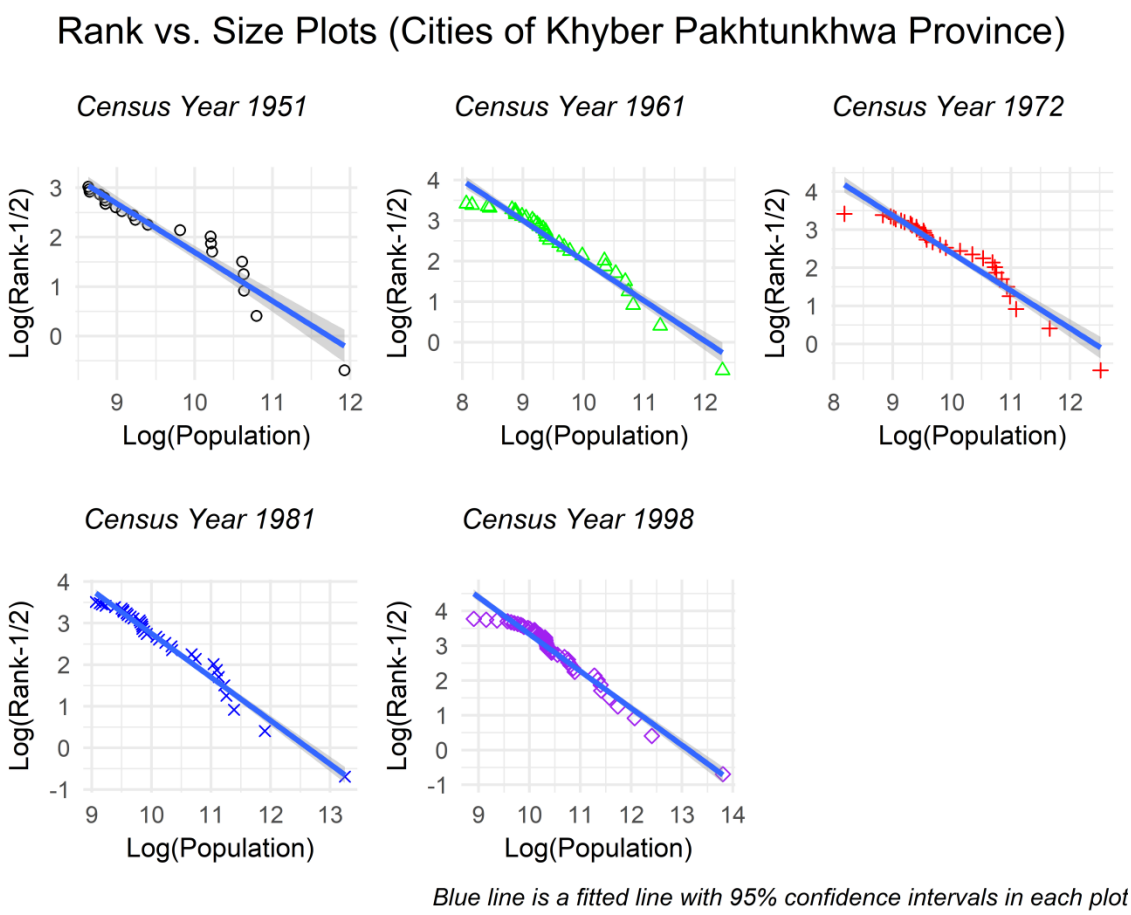


Figure 10 Log (rank-1/2) vs. log (Population) plots for the cities of Khyber Pakhtunkhwa Province

Table 8 Log(Rank-1/2) vs. log(Population) regression results for the cities of Khyber Pakhtunkhwa province

Census Year	Number of cities	*Log (A)*	α	*t-test $\alpha=1$	P-value of t-test
1951	21	11.529	0.983	Do not reject	0.798
1961	31	11.904	0.986	Do not reject	0.804
1972	31	12.250	0.989	Do not reject	0.794
1981	34	13.029	1.046	Do not reject	0.132
1998	44	13.967	1.063	Do not reject	0.065

*95% Confidence interval is used to test hypothesis

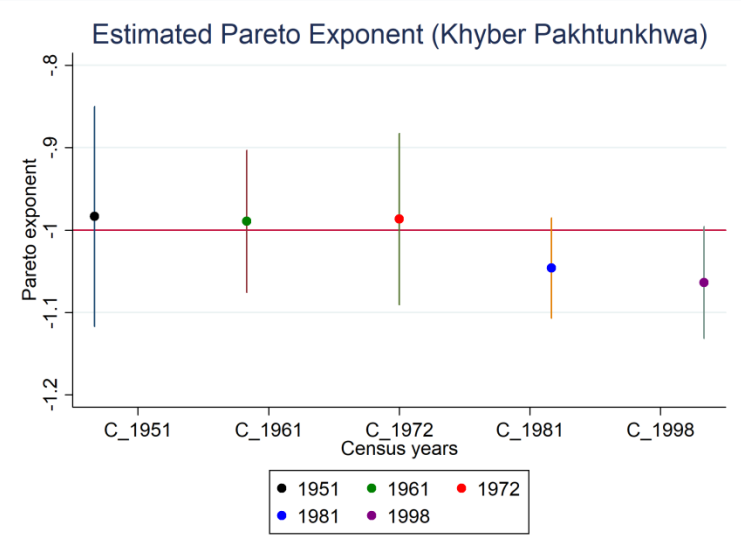
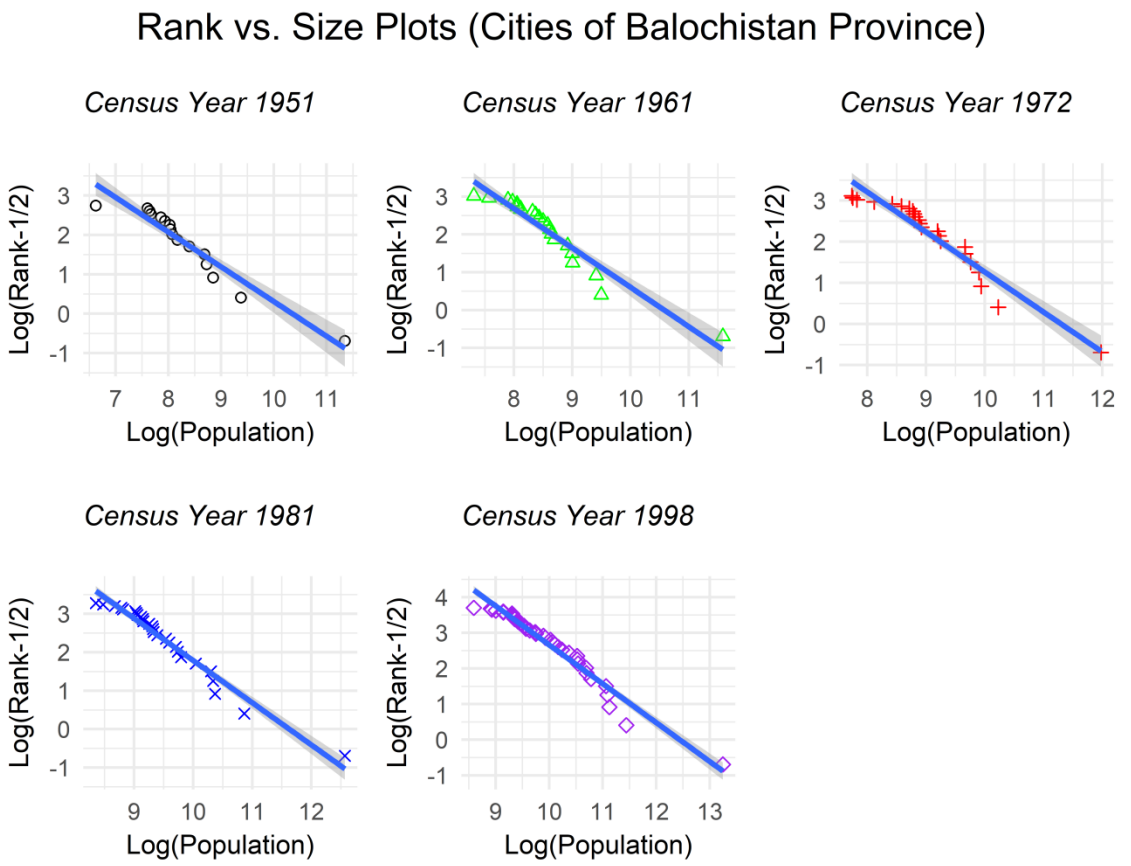


Figure 11 Estimated Pareto exponent (with negative sign as given in Eq. 3) for the cities of Khyber Pakhtunkhwa Province with 95% confidence intervals

Figure (12) shows Zipf's curves, Table (9) reports the results of log(rank-1/2) log(population) regressions and Figure (13) shows the statistical significances of the estimated Pareto exponents for each of the five census years for the cities of Balochistan province. As shown in Figure (13), the estimated values of Pareto exponent for the cities of Balochistan province are statistically equal to 1 in the first three census years of 1951, 1961 and 1972, while are slightly higher than 1 in last two census years. The support to Zipf's law in early census years shows that urban landscape of the province was not changed much during the partition in 1947. However the higher than 1 values of the exponent in later years suggest that the city size distribution has become more even and the size difference between larger and smaller cities has decreased[1]. There are at least two factors which have contributed to this distortion in city size distribution in later years: the lack of employment opportunities and political instability. The major cities in the Baluchistan province are deficient in employment opportunities due to resource deficiency and the lack of consistent development policy for the province by the federal government. This lack of development in the province has created a sense of deprivation in the masses which has

[1] Quetta which is the largest city in Baluchistan province is an exception. Other large cities (i.e., second, third, fourth and so on) are much smaller compared to their predicted size by the Zipf's law.

contributed to the political instability in the province. Because of these reasons, the size of large cities has not increased to the level which is consistent with Zipf's law distribution.



Blue line is a fitted line with 95% confidence intervals in each plot

Figure 12 Log (rank-1/2) vs. log (Population) plots for the cities of Balochistan Province

Table 9 Log(Rank-1/2) vs. log(Population) regression results for the cities of Balochistan province

Census Year	Number of cities	*Log (A)*	α	*t-test $\alpha=1$	P-value of t-test
1951	16	9.117	0.881	Do not reject	0.103
1961	21	10.996	1.039	Do not reject	0.577
1972	23	10.996	0.973	Do not reject	0.663
1981	27	12.762	1.098	Rejected	0.035
1998	41	13.575	1.091	Rejected	0.013

*95% Confidence interval is used to test hypothesis

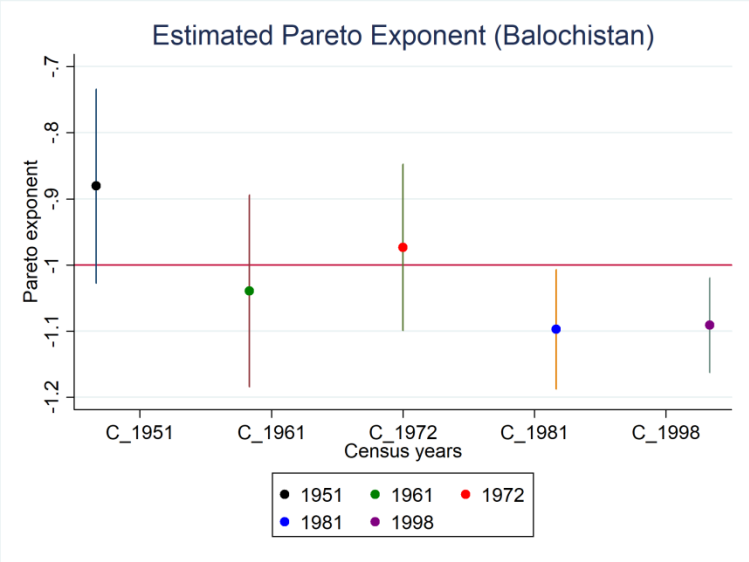


Figure 13 Estimated Pareto exponent (with negative sign as given in Eq. 3) for the cities of Balochistan Province with 95% confidence intervals

In above results, the estimated values of Pareto exponent for the cities within provinces are near to 1 as compared to the estimated values for all cities at national level. Estimated Pareto exponent is 0.90 to 0.94 in different census years for the cities at national level. While for the cities at province-level, these values are between 0.96 to 1.02 for Punjab, 0.94 to 1.06 for Sindh (excluding Karachi), 0.98 to 1.08 for Khyber Pakhtunkhwa, and 0.97[2] to 1.09 for Baluchistan. Overall these findings confirm that city size distributions within provinces are more likely to obey Zipf's law.

Existing literature offers several explanations that why Zipf's law applies to city size distribution. For example, a number of recent studies argue that in industrialized countries, cities concentrate not only a large part of the population but also the economic activity, and the urban structure is an outcome of dynamic interplay between economic activity and growth process of

[2] We didn't consider 0.88 value of the exponent in the year 1951 due to very small sample size.

cities [38, 56-62]. Economic shocks, such as the productivity or innovations shocks, generate skew in city size distribution. When these economic shocks are randomly distributed across cities, the size distribution of cities would be Pareto with Pareto exponent exactly equals to one. In similar vein, Giesen and Südekum [21] argue that if city growth rates are size independent then any random or regional sample from a population of all cities of a country follows the same distribution as the distribution of the population of all cities. Specifically, they find that because the national level city size distribution of Germany is Zipf's law so the any random or regional sample of cities in Germany also follows Zipf's law.

However, our above results are not consistent with the findings of Giesen and Südekum [21]. We find that although Zipf's law does not apply to the city size distribution of Pakistani cities at national level, however the city size distributions at regional level (i.e., the province level) are more likely to be consistent with Zipf's law. Our findings cannot be explained based on economic factors. In Pakistan, the main economic activity is at national level. Karachi is the main port city with major import and export activity. Karachi is linked with other parts of the country through several highways. The main national highway (N-5) is the longest highway in Pakistan. It starts from Karachi, extends through Hyderabad, Moro and Khairpur in Sindh before crossing into Punjab province where it passes through Multan, Sahiwal, Lahore, Gujranwala, Gujrat, Jhelum and Rawalpindi. At Rawalpindi, it turns eastwards and passes through Attock Khurd before crossing the Indus River into Khyber Pakhtunkhwa to continue through Nowshera and Peshawar before entering the Khyber Pass and reaching the border town of Torkham in the FATA. Similarly other parts are also connected through main highways as shown in Fig 14. If economic integration has a role, the city size distribution should be more consistent with Zipf's law at national level. One concern is the previous studies which support economic factors mainly

focus on developed countries. However, Pakistan is a less industrialized country where industrial output accounts for just below 20 percent of total GDP. In this backdrop, we can expect that economic factors have not contributed much to urbanization growth and we explain our results based on social and institutional factors which are likely to have caused the urbanization pattern in Pakistan that is a developing country. This is also evident from the 1998 census where only 12 per cent lifetime migrant said that they moved to cities for employment and 9 per cent for business. While, a largest percentage of migrants moved to cities due to social reasons such as the family or better education and security situation in cities.

Figure 14 Road connectivity in Pakistan (Source: Pakistan property news at http://eproperty.pk/pakistan-national-highways-network-map)

4.3 Coherence property and the city size distribution of provinces

In this section, we explain that the systems of cities within each of the four Pakistani provinces are more coherent in terms of common language, common culture and common institutions as compared to the level of coherence at national level.

Figure (15) shows the languages spoken at national-level and in each of the four provinces of Pakistan. Urdu is national language. Punjabi, Pushto, Sindhi, Balochi and Siraiki are other major regional languages. The most spoken language at national level is Punjabi with 44 percent population speaking it. Comparatively, the percentage of common language at provinces-level is quite higher. For instance, 75.2 percent population speaks Punjabi in Punjab, 73.9 percent population speaks Pushto in Khyber Pakhtunkhwa, 59.73 percent population speaks Sindhi in Sindh and 54.76 percent population speaks Balochi in Balochistan. These statistics show that population is more coherent within provinces in terms of a common language. Since the language plays a vital role in relocation decisions, the population movement within provinces is more likely than across provinces.

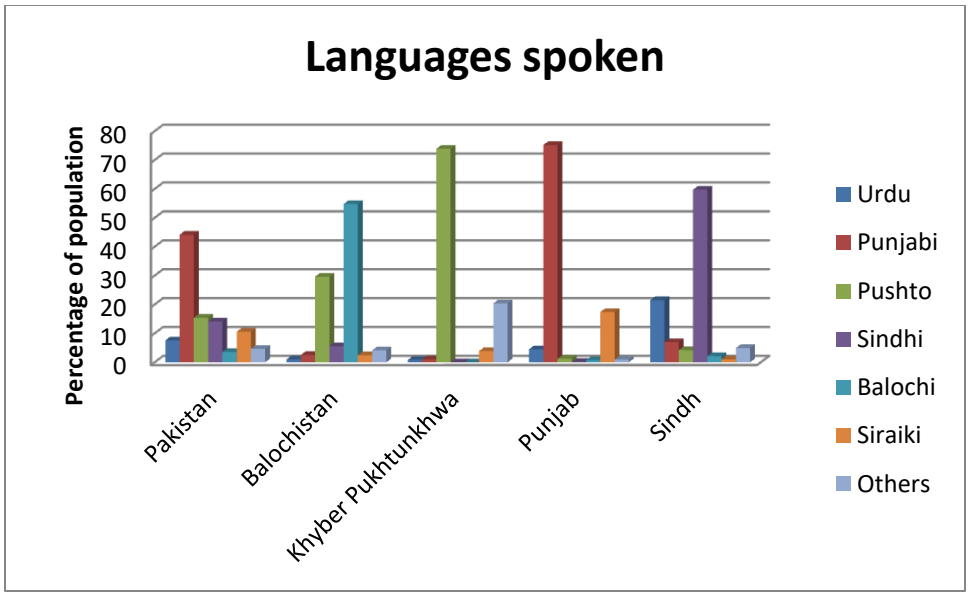


Figure 15 Languages spoken at national and provincial levels (source: 1998 Census as reported by Pakistan Bureau of Statistics at http://www.pbscensus.gov.pk/content/population-and-housing-indicators)

Second, the provinces have more common culture. Pakistan has six major ethnic groups as shown in Figure 16: Punjabis, Seraikis, Sindhis, Pashtuns, Balochis, Muhajirs. These ethnic groups have different cultures in terms of history, dress, cuisine, poetry, music and values.

Punjabis are the largest ethnic group in Pakistan and accounts for 44.7% of the population. Punjabi identity is traditionally geographical, linguistic and cultural and its history dates back to 5th and 4th centuries BC. Punjabis practice different religions including Islam (Majority are Muslims in Pakistani Punjab[3]), Hinduism, Sikhism and Christianity. Punjabi culture has strong influence of kinship-based social networks where a person faces a strong pressure to pool and share resources, such as income, personal connections and political influence, among kins. The culture comprises of spirituality, poetry, music, cuisine, language, history, and values.

[3] Punjab was distributed into two parts (Pakistani Punjabi and Indian Punjab) at the time of partition in 1947.

Pushtuns, also referred to as Pathans, are the second largest ethnic group in Pakistan. Pushtuns largely adhere to Pashtunwali code which defines the culture of the Pashtun and involves a self-governing tribal system that controls all aspects of the group.

Sindhis are third largest ethnic group. Sindhi culture has a very glorious past. Archaeological researchers have shown that the roots of Sindhi culture in terms of agricultural practices, traditional arts and crafts, customs and tradition etc. dates back to the mature Indus valley civilization of the third millennium B.C.

Seraikis are fourth largest group in Pakistan and are settled in southern part of Punjab. This is the second largest community in Punjab province, just after Punjabis. Seraiki language is considered as one of the dialects of Punjabi language, and therefore Seraiki culture has been greatly influenced by the Punjabi culture because of their close association.

Balochis, the fifth largest ethnic community, are a group of tribes speaking the Balochi language. Balochis are divided into two groups, the Sulaimani and the Makrani, separated from each other by a compact block of Brahui tribes.

These ethnicities have strong effect on population distribution in Pakistan. Punjabis and Siraikis are largely live in Punjab province, Pushtuns in Khyber Pakhtunkhwa, Sindhis in Sindh and Balochis in Baluchistan province. Further, the ethnic cultures strongly influence people's decisions such as whether and where to migrate. Due to kinship-based Punjabi and tribe-based Pushtun and Baloch cultures, individuals prefer to move within the same province where the culture is largely same. Similarly, when individuals migrate from rural to urban area, they chose destinations where it is easy to keep contact with left behinds.

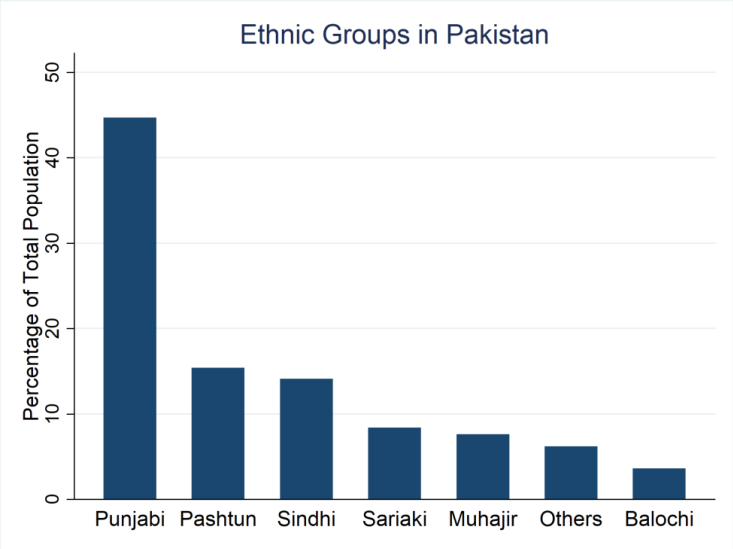


Figure 16 Proportion of major ethnic groups in total population of Pakistan (Source: South Asia-Pakistan in CIA The World Factbook)

Similarly, provinces have more common rules. Pakistan has mainly two tier political system: federal government and provincial governments. Federal government is responsible for national-level issues such as defense and foreign diplomacy. Provinces enjoy a lot of autonomy, which has been further increased through the eighteenth amendment in the constitution of the country. The resource distribution decisions within provinces largely lie with provincial governments. Further, provincial jobs are allocated to individuals from the same province. For instance, each province has its own provincial services commission who is responsible for hiring individuals for bureaucratic jobs. Similarly, jobs in provincial education and health departments also remain in provincial domain. Moreover, students are admitted in government educational institutions (including provincial universities) based on the residence province. Domicile certificate is usually used to ascertain the residence province of an individual.

5. Conclusion

In this paper, we examine the Zipf's law for the size distribution of Pakistani cities using the five census years' data from 1951 to 1998. We observe that Zipf's law does not hold in any of the five census years at national level. Next, we distribute the main sample into four province-wise subsamples and observe that Zipf's law is more likely to hold for the city-size distribution at province-level. We attribute these findings to the coherence property of the urban system because in Pakistan the urban systems within provinces are more coherent in terms of common language, common culture and common rules as compared to the urban system as a whole at national level.

References

[1] M. Jefferson, The Law of the Primate City, Geographical Review, 29 (1939) 226-232.
[2] A.F. Ades, E.L. Glaeser, Trade and circuses: explaining urban giants, The Quarterly Journal of Economics, 110 (1995) 195-227.
[3] V. Henderson, Urbanization in developing countries, The World Bank Research Observer, 17 (2002) 89-112.
[4] Z.I. Dimitrova, M. Ausloos, Primacy analysis in the system of Bulgarian cities, Open Physics, 13 (2015) 218-225.
[5] J. Eeckhout, Gibrat's Law for (All) Cities, The American Economic Review, 94 (2004) 1429-1451.
[6] K. Giesen, A. Zimmermann, J. Suedekum, The size distribution across all cities–double Pareto lognormal strikes, Journal of Urban Economics, 68 (2010) 129-137.
[7] E. Gómez-Déniz, E. Calderín-Ojeda, On the use of the Pareto ArcTan distribution for describing city size in Australia and New Zealand, Physica A: Statistical Mechanics and its Applications, 436 (2015) 821-832.
[8] R. González-Val, A. Ramos, F. Sanz-Gracia, M. Vera-Cabello, Size distributions for all cities: Which one is best?, Papers in Regional Science, 94 (2015) 177-196.
[9] Y.M. Ioannides, S. Skouras, US city size distribution: Robustly Pareto, but only in the tail, Journal of Urban Economics, 73 (2013) 18-29.
[10] J. Luckstead, S. Devadoss, Pareto tails and lognormal body of US cities size distribution, Physica A: Statistical Mechanics and its Applications, 465 (2017) 573-578.
[11] J. Luckstead, S. Devadoss, D. Danforth, The size distributions of all Indian cities, Physica A: Statistical Mechanics and its Applications, 474 (2017) 237-249.
[12] L.C. Malacarne, R. Mendes, E.K. Lenzi, q-Exponential distribution in urban agglomeration, Physical Review E, 65 (2001) 017106.
[13] J.B. Parr, K. Suzuki, Settlement populations and the lognormal distribution, Urban Studies, 10 (1973) 335-352.
[14] W.J. Reed, M. Jorgensen, The Double Pareto-Lognormal Distribution—A New Parametric Model for Size Distributions, Communications in Statistics - Theory and Methods, 33 (2004) 1733-1753.
[15] J.M. Sarabia, F. Prieto, The Pareto-positive stable distribution: A new descriptive model for city size data, Physica A: Statistical Mechanics and its Applications, 388 (2009) 4179-4191.
[16] C.A. Vallabados, S.A. Arumugam, An Evaluation of Pareto, Lognormal and PPS Distributions: The Size Distribution of Cities in Kerala, India, Journal of Modern Applied Statistical Methods, 15 (2016) 41.
[17] P. Krugman, Confronting the Mystery of Urban Hierarchy, Journal of the Japanese and International Economies, 10 (1996) 399-418.
[18] J. Eaton, Z. Eckstein, Cities and growth: Theory and evidence from France and Japan, Regional Science and Urban Economics, 27 (1997) 443-474.
[19] Y.M. Ioannides, H.G. Overman, Zipf's law for cities: an empirical examination, Regional Science and Urban Economics, 33 (2003) 127-137.
[20] L. Gligor, M. Gligor, The fractal city theory revisited: new empirical evidence from the distribution of Romanian cities and towns, Nonlinear dynamics, psychology, and life sciences, 12 (2008) 15-28.
[21] K. Giesen, J. Südekum, Zipf's law for cities in the regions and the country, Journal of Economic Geography, 11 (2011) 667-686.
[22] B.J. Berry, A. Okulicz-Kozaryn, The city size distribution debate: Resolution for US urban regions and megalopolitan areas, Cities, 29 (2012) S17-S23.
[23] J. Luckstead, S. Devadoss, A comparison of city size distributions for China and India from 1950 to 2010, Economics Letters, 124 (2014) 290-295.

[24] E. Ezzahid, O. ElHamdani, Zipf'S Law In The Case Of Moroccan Cities, Review of Urban & Regional Development Studies, 27 (2015) 118-133.
[25] G. Anderson, Y. Ge, The size distribution of Chinese cities, Regional Science and Urban Economics, 35 (2005) 756-776.
[26] V. Nitsch, Zipf zipped, Journal of Urban Economics, 57 (2005) 86-100.
[27] K.T. Soo, Zipf's Law for cities: a cross-country investigation, Regional Science and Urban Economics, 35 (2005) 239-263.
[28] M. Bee, M. Riccaboni, S. Schiavo, The size distribution of US cities: Not Pareto, even in the tail, Economics Letters, 120 (2013) 232-237.
[29] A. Lalanne, Zipf's Law and Canadian Urban Growth, Urban Studies, 51 (2014) 1725-1740.
[30] J. Luckstead, S. Devadoss, Do the world's largest cities follow Zipf's and Gibrat's laws?, Economics Letters, 125 (2014) 182-186.
[31] K.T. Soo, Zipf's Law and Urban Growth in Malaysia, Urban Studies, 44 (2007) 1-14.
[32] E. Deliktas, A.Ö. Önder, M. Karadag, The size distribution of cities and determinants of city growth in Turkey, European Planning Studies, 21 (2013) 251-263.
[33] H.E. Duran, S.P. Ozkan, Trade Openness, Urban Concentration And City-Size Growth In Turkey, Regional Science Inquiry, 7 (2015) 35-46.
[34] E. Pérez-Campuzano, L. Guzmán-Vargas, F. Angulo-Brown, Distributions of city sizes in Mexico during the 20th century, Chaos, Solitons & Fractals, 73 (2015) 64-70.
[35] S. Arshad, S. Hu, B.N. Ashraf, Zipf's law and city size distribution: A survey of the literature and future research agenda, Physica A: Statistical Mechanics and its Applications, 492 (2018) 75-92.
[36] The World Bank, World Development Indicators Database, http://data.worldbank.org/data-catalog/world-development-indicators (2014).
[37] G. Fazio, M. Modica, Pareto or Log-Normal? Best Fit and Truncation in the Distribution of All Cities, Journal of Regional Science, 55 (2015) 736-756.
[38] X. Gabaix, Zipf's law for cities: an explanation, Quarterly journal of Economics, (1999) 739-767.
[39] M. Levy, Gibrat's Law for (All) Cities: Comment, The American Economic Review, 99 (2009) 1672-1675.
[40] Y. Malevergne, V. Pisarenko, D. Sornette, Gibrat's law for cities: uniformly most powerful unbiased test of the Pareto against the lognormal, Swiss Finance Institute Research Paper, (2009).
[41] Y. Malevergne, V. Pisarenko, D. Sornette, Testing the Pareto against the lognormal distributions with the uniformly most powerful unbiased test applied to the distribution of cities, Physical Review E, 83 (2011) 036111.
[42] Y. Chen, Y. Zhou, Scaling laws and indications of self-organized criticality in urban systems, Chaos, Solitons & Fractals, 35 (2008) 85-98.
[43] K. Gangopadhyay, B. Basu, City size distributions for India and China, Physica A: Statistical Mechanics and its Applications, 388 (2009) 2682-2688.
[44] K. Gangopadhyay, B. Basu, Evolution of Zipf's Law for Indian Urban Agglomerations Vis-à-Vis Chinese Urban Agglomerations, in: Econophysics of Systemic Risk and Network Dynamics, Springer, 2013, pp. 119-129.
[45] H. Li, Y.D. Wei, Y. Ning, Spatial and Temporal Evolution of Urban Systems in China during Rapid Urbanization, Sustainability, 8 (2016) 651.
[46] S. Li, D. Sui, Pareto's law and sample size: a case study of China's urban system 1984–2008, GeoJournal, 78 (2013) 615-626.
[47] S. Song, K.H. Zhang, Urbanisation and city size distribution in China, Urban Studies, 39 (2002) 2317-2327.
[48] X. Ye, Y. Xie, Re-examination of Zipf's law and urban dynamic in China: a regional approach, Ann Reg Sci, 49 (2012) 135-156.

[49] W. Ziqin, Zipf Law Analysis of Urban Scale in China, Asian Journal of Social Science Studies, 1 (2016) 53.
[50] C.A. Ignazzi, The Brazilian Urban System: the trajectories of Brazilian cities between general dynamics and specific peculiarities, Cybergeo: European Journal of Geography, (2015) document 754.
[51] V.J. Matlaba, M.J. Holmes, P. McCann, J. Poot, A Century of the Evolution of the Urban System in Brazil, Review of Urban & Regional Development Studies, 25 (2013) 129-151.
[52] N.J. Moura, M.B. Ribeiro, Zipf law for Brazilian cities, Physica A: Statistical Mechanics and its Applications, 367 (2006) 441-448.
[53] F. Guerin-Pace, Rank-Size Distribution and the Process of Urban Growth, Urban Studies, 32 (1995) 551-562.
[54] G. Peng, Zipf's law for Chinese cities: Rolling sample regressions, Physica A: Statistical Mechanics and its Applications, 389 (2010) 3804-3813.
[55] M. Cristelli, M. Batty, L. Pietronero, There is more than a power law in Zipf, Scientific reports, 2 (2012).
[56] J.-C. Córdoba, On the distribution of city sizes, Journal of Urban Economics, 63 (2008) 177-197.
[57] J.C. Córdoba, A generalized Gibrat's law, International Economic Review, 49 (2008) 1463-1468.
[58] G. Duranton, Some foundations for Zipf's law: Product proliferation and local spillovers, Regional science and urban Economics, 36 (2006) 542-563.
[59] G. Duranton, Urban evolutions: The fast, the slow, and the still, The American Economic Review, 97 (2007) 197-221.
[60] W.T. Hsu, Central place theory and city size distribution, The Economic Journal, 122 (2012) 903-932.
[61] S. Lee, Q. Li, Uneven landscapes and city size distributions, Journal of Urban Economics, 78 (2013) 19-29.
[62] E. Rossi-Hansberg, M.L. Wright, Urban structure and growth, The Review of Economic Studies, 74 (2007) 597-624.
[63] F. Auerbach, Das gesetz der bevölkerungskonzentration, Petermanns Geographische Mitteilungen, 59 (1913) 74-76.
[64] H.W. Singer, The "Courbe des Populations." A Parallel to Pareto's Law, The Economic Journal, 46 (1936) 254-263.
[65] G.K. Zipf, Human behavior and the principle of least effort, Reading, MA: Addison-Wesley (1949).
[66] X. Gabaix, Y.M. Ioannides, The evolution of city size distributions, Handbook of regional and urban economics, 4 (2004) 2341-2378.
[67] Y. Konishi, Y. Nishiyama, Hypothesis testing in rank-size rule regression, Mathematics and Computers in Simulation, 79 (2009) 2869-2878.
[68] Y. Nishiyama, S. Osada, Y. Sato, OLS estimation and the *t* test revisited in rank-size rule regression, Journal of Regional Science, 48 (2008) 691-716.
[69] X. Gabaix, R. Ibragimov, Rank − 1 / 2: A Simple Way to Improve the OLS Estimation of Tail Exponents, Journal of Business & Economic Statistics, 29 (2011) 24-39.
[70] E. Uriel, Hypothesis testing in the multiple regression model, Universidad de Valencia: Department of economics, (2013).
[71] P. Cheshire, Trends in sizes and structures of urban areas, Handbook of regional and urban economics, 3 (1999) 1339-1373.
[72] B. Jiang, T. Jia, Zipf's law for all the natural cities in the United States: a geospatial perspective, International Journal of Geographical Information Science, 25 (2011) 1269-1281.
[73] B. Jiang, X. Liu, Scaling of geographic space from the perspective of city and field blocks and using volunteered geographic information, International Journal of Geographical Information Science, 26 (2012) 215-229.

[74] B. Jiang, J. Yin, Q. Liu, Zipf's law for all the natural cities around the world, International Journal of Geographical Information Science, 29 (2015) 498-522.
[75] K.T. Rosen, M. Resnick, The size distribution of cities: An examination of the Pareto law and primacy, Journal of Urban Economics, 8 (1980) 165-186.
[76] H.D. Rozenfeld, D. Rybski, X. Gabaix, H.A. Makse, The Area and Population of Cities: New Insights from a Different Perspective on Cities, The American Economic Review, 101 (2011) 2205-2225.
[77] K. Schmidheiny, J. Suedekum, The pan-European population distribution across consistently defined functional urban areas, Economics Letters, (2015).
[78] P. Veneri, City size distribution across the OECD: Does the definition of cities matter?, Computers, Environment and Urban Systems, 59 (2016) 86-94.
[79] G. Arif, S. Hamid, Urbanization, city growth and quality of life in Pakistan, European Journal of Social Sciences, 10 (2009) 196-215.
[80] M. Kugelman, Urbanisation in Pakistan: causes and consequences, Washington DC: NOREF Norwegian Peacebuilding Resource Center, (2013).
[81] I. Ara, A. Zaman, Asian urbanization in the new millennium: Pakistan chapter, unpublished paper written for an Asian Urban Information Centre publication for Kobe, (2002).
[82] H.A. Simon, On a class of skew distribution functions, Biometrika, (1955) 425-440.