JOURNAL OF URBAN ECONOMICS 8, 165–186 (1980)

The Size Distribution of Cities: An Examination of the Pareto Law and Primacy

KENNETH T. ROSEN AND MITCHEL RESNICK

Center for Urban and Real Estate Economics, School of Business Administration, University of California, Berkeley, California 94720

Received September 20, 1978; revised December 1, 1978

This paper examines the Pareto and primacy measures of the size distribution of cities. The mean Pareto exponent for a sample of 44 countries is 1.136, somewhat greater than the exponent of one implied by the rank-size rule. We find that value of the Pareto exponent is quite sensitive to the definition of the city and the choice of city sample size. The significance of non-linear terms in variants of the Pareto distribution also indicate that the rank-size rule is only a first approximation to a complete characterization of the size distribution of cities within a country. The relatively low correlation between primacy and Pareto measures confirms the need for a variety of measures of city size distributions. This paper also suggests that large cities are growing faster than small cities in most of the countries in our sample. This is indicated by the positive coefficient on the first non-linear term introduced into the Pareto equation. Finally, variations in the Pareto exponent and measures of primacy are partly explained by economic, demographic, and geographic factors.

1. INTRODUCTION

The literature on the size distribution of cities within an urban hierarchy is extensive. Much of this literature has focused on applying the Pareto distribution, which appears to be a close approximation to many size distributions in economics, to the distribution of city sizes within a country. (See for example, Berry [4], Beckman [1], Vining [9], Parr and Suzuki [7], Mills [6]). Many findings and policy discussions focus on a special case of the Pareto distribution known in the urban literature as the "rank-size rule." The "rank-size rule" states that for the cities within a given country the product of a city's rank and its population is approximately equal to a constant. In terms of the Pareto distribution, the rank-size rule implies a Pareto exponent of unity.

While the empirical regularity associated with the rank-size rule has been tested for a number of countries, there has been no recent systematic test of the relationship for a sample of countries encompassing both developing and developed nations. Such a test is important, given the extensive policy debate over the degree to which large urban agglomerations are becoming so dominant in the urban hierarchy as to create a sub-optimal use of economic resources in these countries.

0094-1190/80/050165-22$02.00/0

This paper provides new empirical insights into the Pareto distribution and primacy debates. First, we examine the size distribution of cities for a large number of countries using the Pareto distribution and alternative measures of primacy. We also compare the primacy concept to the "Pareto" measure of the size distribution. Second, we examine the sensitivity of the results to alternative assumptions concerning the number of cities included in the sample and the definition of a "city." Third, we examine a more elaborate "non-Pareto" behavior which provides additional insight into the size distribution of cities. Finally, variations in the Pareto coefficient and in the extent of primacy across countries are explained by economic, demographic, and geographic factors.

2. THE PARETO DISTRIBUTION AND MEASURES OF PRIMACY

The Pareto distribution for city-size distributions is represented in mathematical form in (1).

$$R = AS^{-a}, \qquad (1)$$

where

R = number of cities with population S or more,

A = constant

S = population of city,

a = Pareto exponent.

The rank-size rule is the special case of this distribution when $a = 1$. Equation (1) was estimated in double logarithmic form for 44 countries. For most countries this relationship was estimated for the 50 largest cities in the country using 1970 Census data. In countries in which the 50 largest cities had more than 100,000 people, all cities over 100,000 were included in the data sample. Issues of data definition are discussed in more detail in Section III and in the Appendix.

Table 1, Column (1) presents the Pareto exponent values for 44 countries, with the rank of the country in brackets. The Pareto values range from 0.809 (Morocco) to 1.963 (Australia). Figures 1 and 2 show the plot of rank versus size for the United States and the Soviet Union. The mean Pareto value is 1.136 and the standard deviation is 0.196. Almost three-fourths (32 of 44) of the countries have exponents greater than unity. This indicates that populations in most countries are more evenly distributed than would be predicted by the rank-size rule (Pareto exponent = 1). This relationship between higher Pareto exponent and more evenly distributed population can be seen by considering the limiting value of $a \to \infty$, in which case all cities would have the same population.

TABLE 1
Pareto Distribution and Measures of Primacy

	Pareto Exponent (1)	Primacy I (2)	Primacy II (3)
Argentina	0.933(5)	0.5288(17)	0.2271(23)
Australia	1.963(44)	0.5291(16)	0.1442(38)
Austria	0.875(3)	0.6990(3)	0.5023(1)
Brazil	1.153(29)	0.4395(28)	0.2189(25)
Canada	1.132(25)	0.3553(39)	0.1292(41)
Colombia	0.847(2)	0.4799(24)	0.2932(10)
Czechoslavakia	1.107(21)	0.5075(20)	0.2580(15)
Denmark	1.374(41)	0.5559(8)	0.2232(24)
Ethiopia	0.970(9)	0.6863(5)	0.4426(4)
Finland	1.084(15)	0.5089(19)	0.2362(21)
France	1.325(40)	0.5511(10)	0.2449(18)
E. Germany	1.125(22)	0.3959(31)	0.1833(27)
W. Germany	1.171(30)	0.3141(43)	0.1135(42)
Ghana	1.104(19)	0.5493(12)	0.2973(8)
Greece	1.138(27)	0.5316(14)	0.2634(13)
Hungary	1.092(17)	0.7646(1)	0.4543(3)
India	1.204(34)	0.3622(35)	0.1550(37)
Indonesia	0.967(7)	0.5311(15)	0.2936(9)
Iran	0.993(11)	0.6430(7)	0.3589(6)
Israel	0.983(10)	0.3501(40)	0.1650(34)
Italy	1.046(14)	0.3614(37)	0.1769(30)
Japan	1.289(39)	0.5048(21)	0.2393(19)
Malaysia	0.968(8)	0.3707(34)	0.1652(33)
Mexico	1.153(28)	0.4885(22)	0.2310(22)
Morocco	0.809(1)	0.5418(13)	0.3105(7)
Netherlands	1.266(37)	0.3278(41)	0.1361(40)
Nigeria	1.537(43)	0.3142(42)	0.0974(44)
Norway	1.265(36)	0.5502(11)	0.2459(17)
Philippines	1.253(35)	0.4292(30)	0.1830(28)
Poland	1.127(23)	0.3589(38)	0.1429(39)
Romania	1.085(16)	0.6596(6)	0.2821(12)
S. Africa	0.997(12)	0.2329(44)	0.1047(43)
Spain	1.133(26)	0.4787(25)	0.2372(20)
Sri Lanka	1.13(24)	0.5540(9)	0.2621(14)
Sweden	1.41(42)	0.4351(29)	0.1711(32)
Switzerland	1.095(18)	0.3811(33)	0.1822(29)
Thailand	0.961(6)	0.6907(4)	0.4775(2)
Turkey	1.077(14)	0.4881(23)	0.2580(16)
U.K.	1.178(31)	0.7070(2)	0.3621(5)
U.S.S.R.	1.278(38)	0.4724(26)	0.1560(36)
U.S.A.	1.184(32)	0.4502(27)	0.1945(26)
Venezuela	1.106(20)	0.3919(32)	0.1715(31)
Yugoslavia	1.186(33)	0.3618(36)	0.1607(35)
Zaire	0.930(4)	0.5176(18)	0.2870(11)
Mean	1.136	0.4852	0.2372
Standard deviation	(0.196)	(0.1219)	(0.0980)

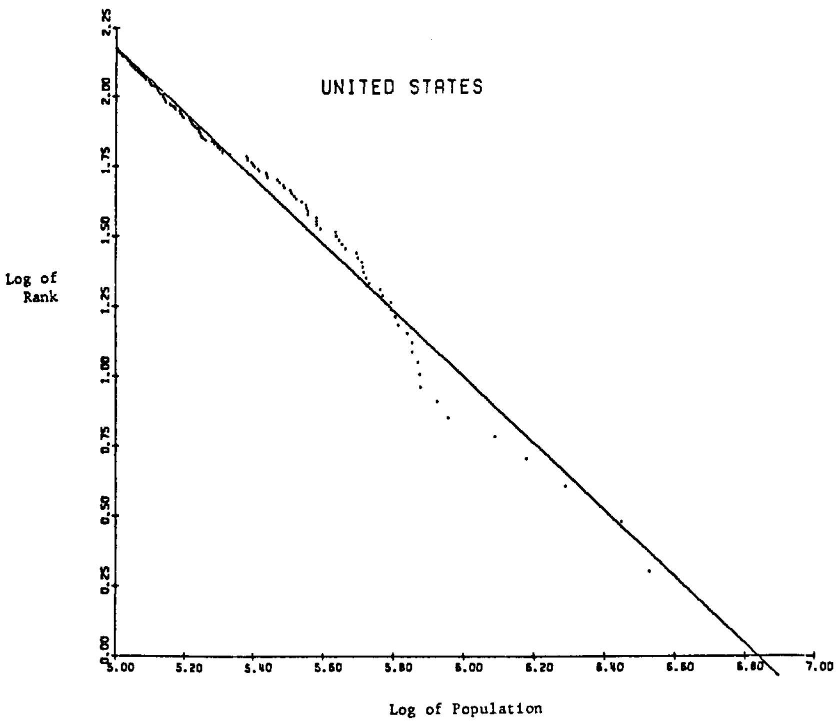

FIGURE 1

The mean Pareto value is inflated by Australia's 1.963, which falls more than four standard deviations from the mean. Australia is clearly an exceptional case. Yet even if we exclude Australia from our sample, Pareto values tend to be greater than unity. Ten other countries have exponents exceeding 1.20, while none have exponents below 0.80. This indicates that the rank-size rule should be reconsidered. The substantial variance across countries in the Pareto exponent also suggests that urban economists and geographers need to derive a theoretical framework which can explain these variations.

While urban economists have focused mainly on the Pareto measure of the size distribution of cities, development economists have focused on measures of primacy as their characterization of the urban system. Primacy can be defined in various ways, such as the ratio of the largest city to the sum of the first two, three, four, or more cities. Generally, however, a measure of primacy is chosen to reflect the extent to which the largest city or cities dominate the country's urban hierarchy. In our study we used two measures of primacy: the ratio of the largest city to the sum of the top five

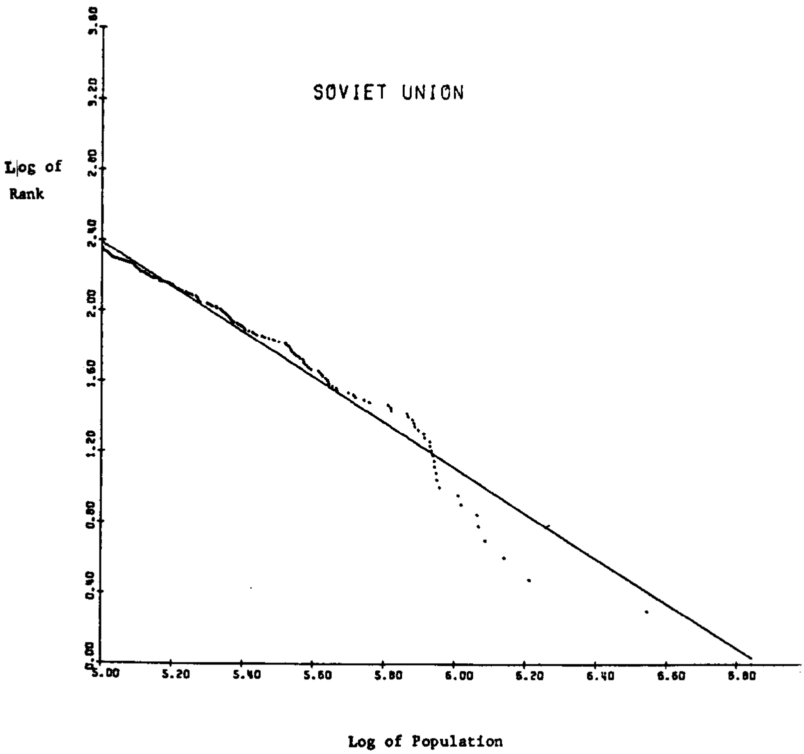

FIGURE 2

cities (Primacy I) and the ratio of the largest to the sum of the top 50 cities (Primacy II). The second measure should provide a better reflection of how the top city compares to the distribution as a whole, rather than just to the rest of the upper tail.

Table 1, columns (2) and (3), shows the two measures along with the primacy ranking of the country. The two measures of primacy are strongly correlated (.922). There are, however, large variations in each measure from country to country. Budapest represents 76% of the people in Hungary's top five cities, and 45% of its top 50 cities. On the other hand, New York contains 45% of the people in top five U.S.A. cities, and only 19% of the population in top 50 cities.

In comparing the Pareto and primacy measures, we expect Pareto exponents to have a negative correlation with measures of primacy, since higher Pareto exponents correspond to more evenly distributed population. We also expect the Pareto exponents to correlate more strongly with Primacy II than Primacy I, since the second measure takes a greater portion of the overall distribution into account. Our results confirmed both hypotheses: the correlation coefficient between the Pareto exponents and Primacy I is -0.3169, and the correlation coefficient using Primacy II is -0.5195.

Our results also confirm that Pareto exponents are a better reflection of the overall city-size distribution than are primacy measures. The case of Australia illustrates this point. Neither measure of primacy suggests that Australia is exceptional, yet Australia's overall distribution is unusual, as can be seen by the difference between the two primacy measures: Australia ranks 16th (out of 44) using the first measure of primacy, and 39th using the second measure. This change in ranking shows that the distribution of population in the intermediate-sized cities (ranked 6th through 50th) must be unusual. The Pareto exponent for Australia reflects this exceptional feature (Australia's exponent of 1.963 is 0.4 larger than for any other country). Thus the single Pareto exponent exposes a feature which both primacy measures together were needed to detect.

The same phenomenon can be seen with many other countries. In general, countries which rank significantly higher using the first measure of primacy (such as Australia, Denmark, France, and U.S.S.R.) must have relatively large intermediate-sized cities (ranked 6th through 50th) to pull down primacy when the second measure is used. This situation corresponds to a relatively uniform population distribution and thus the Pareto exponents for these countries are generally greater than the mean. On the other hand, countries which rank significantly higher using the second measure of primacy (such as Columbia, Italy, Turkey, and Zaire) have relatively small intermediate-sized cities and Pareto exponents are less than the mean.

3. SENSITIVITY OF RESULTS TO DATA DEFINITIONS

A. Definition of "city"

In any study dealing with systems of cities, the question of the appropriate urban unit must be considered. Whether a study uses urban places, legal cities, or urban agglomerations may affect the value of the observed Pareto exponent and the closeness of the fit. The cities in a given country might fit the rank-size rule merely by virtue of the definition chosen.

For size distribution studies, the entire metropolitan area is the most desirable choice for an urban unit as it represents an integrated economic unit. Since many workers and consumers in a city often reside in the surrounding suburbs, it seems reasonable to include these areas in the definition of the city. However, in our cross-country comparisons, we are constrained by the availability of data. Many countries do not provide populations for entire metropolitan areas; therefore, we used "city proper"

TABLE 2
Definition of City and the Pareto Exponent

Country	Pareto exponent using city proper	Pareto exponent using metro pop	% Difference
Brazil	1.153	0.820	– 28.9
France	1.325	1.104	– 16.7
India	1.204	1.125	– 6.6
Italy	1.064	0.958	– 10.0
Mexico	1.153	0.963	– 16.5
United States	1.184	1.000	– 15.5

data, which were available for all countries (see Appendix for sources and years of data). Even using "city proper" figures for all countries, we must use caution in comparing countries since the definition of "city proper" may vary slightly from country to country.

To test the effect of the city definition choice on the resulting Pareto exponent, we selected six countries for which both city proper and urban agglomeration data were available. Table 2 shows the Pareto exponents for these six countries, using each of the two definitions. In each case, the Pareto exponent dropped as city limits expanded. This result was expected since including suburbs causes a greater percentage change for large city populations than for small city populations due to the spillover of population of large cities to suburban areas. Thus population is less evenly distributed using this definition and the value of the Pareto exponent decreases. In the six countries chosen, the exponent decreased anywhere from 6.6 to 28.9%. Although this range is large, most of the changes clustered around 15%. It is also interesting to note that the United States Pareto exponent using metropolitan population is precisely 1.

B. Sample Size

A second question concerns the number of cities used for each country. If the data fit a Pareto distribution perfectly, then the number of cities in the sample is unimportant; it takes only two points to determine a Pareto curve. But since deviations from the Pareto curve occur, our choice of sample size becomes important.

There are two logical ways to define sample size—use a fixed number of cities in each sample, or use a threshold population in each sample. For most countries, we chose the first method, so that each sample consisted of the 50 largest cities in the country. By using a fixed number of cities, each sample is likely to represent approximately the same portion of the overall city-size distribution. If a threshold population had been used to determine sample size, the sample might have included only the upper tail of the distribution for some countries while including the entire distribution for others.

TABLE 3

Sample Size and the Pareto Exponent: Countries with more than 50 cities larger than 100,000

Country	# cities	(1) Pareto exponent[a]	(2) Exponent using top 50 cities	% difference ((1) − (2))/(2)	curvature (See Section IV)
Brazil	91	1.153	1.079	+ 6.4	up
W. Germany	59	1.171	1.186	− 1.3	down
India	149	1.204	1.190	+ 1.2	up
Japan	138	1.289	1.174	+ 8.9	up
U.S.S.R.	225	1.278	1.528	− 19.6	down
U.S.A.	151	1.184	1.310	− 10.6	down

[a] Using all cities more than 100,000.

We made an exception to this sample size rule when considering countries having more than 50 cities with populations exceeding 100,000. For these six countries, we believed that 50 cities did not represent a sufficient portion of the distribution, so we included in the samples all cities with populations greater than 100,000. Table 3 shows how the calculated Pareto exponent is affected by altering the sample size definition. The magnitude of change varies widely from a small rise to a nearly 20% fall in the Pareto exponent with the direction of these changes consistent with the non-Pareto curvature of the rank-size graphs (see Section IV). These results indicate the importance of the choice of sample size.

Closely related to the question of sample size is a more basic question: From what "pool" of cities should each sample be chosen? That is, what constitutes an appropriate "system" of cities? In this study, we use political boundaries to separate our system of cities. Different boundaries might have greater economic rationale, however, especially where political boundaries are open to free exchange of goods and people. A country is less likely to develop a large primate city if it has open borders with countries with dominating primate cities. To achieve consistent results among our samples, we would need to consider more "economically rational systems of cities"—for example perhaps the Common Market countries should be considered as a single entity. Redefining the system of cities was beyond the scope of this study.

4. NON-PARETO BEHAVIOR OF CITY-SIZE DISTRIBUTION

Beyond the question of how closely city-size data obey the rank-size rule is the more fundamental question of how well these populations fit the general Pareto distribution. Many empirical studies of city sizes have focused on the deviations of observed Pareto exponents away from 1.0, taking for granted that the Pareto was an accurate description of the samples being studied. Several recent studies (Vining [9] Parr and Suzuki [7]), however, have discussed reasons for non-Pareto behavior. In this section, we report on deviations from the Pareto form, regularities in these deviations, and the implications of regular deviations.

In order to test for "non-Pareto behavior" we added non-linear terms to the basic logarithmic version of (1), giving (2) and (3),

$$\log R = a' + b' \log P + c' (\log P)^2 \tag{2}$$

$$\log R = a'' + b'' \log P + c'' (\log P)^2 + d'' (\log P)^3 \tag{3}$$

Table 4 shows the results of these regressions. The standard errors of the coefficients are shown in brackets. The quadratic and cubic terms are quite significant in most cases.

More interesting than the significance of these extra terms is the direction of curvature which they indicate. b' is an indicator of curvature, with $b' > 0$ indicating upward concavity and $b' < 0$ downward concavity. Previous studies have shown that rank-size data (for cities or firms) when graphed on double log paper generally tend to exhibit a downward concavity (Vining [9]). In this study, however, more than two-thirds (30 of 44) of the countries showed an *upward* concavity.

This unexpected result calls for a re-evaluation of the theories which have been advanced to explain non-Pareto behavior. It has been suggested that the curvature of the rank-size plots results from the violation of Gibrat's law of proportionate effect, which states that percentage rates of growth (of cities or firms) are independent of size. Gibrat's law was one of the fundamental assumptions used by Simon in generating a Pareto distribution. The previously observed downward concavity has been explained as a negative correlation between growth rates and size, thus producing more cities in the intermediate size classes than would be predicted by a Pareto distribution. However, if an upward concavity is characteristic of most countries, as is suggested by our results, it seems likely that the growth rates of cities are positively correlated to size in these countries. This may reflect scale economies. In many countries, the highest-ranking cities, taking advantage of scale economies, have grown most quickly, causing many writers to worry about "over-urbanization." This is especially true of developing countries and certain mature industrial countries such as France. With such a rapid growth of the largest cities, we expect rank-size graphs to take on an upward concavity. Downward concavity is exhibited by only a few mature industrial countries (Canada and the United States included).

TABLE 4

City Size Distributions—1970

	LPOP	*LPOP2*	*LPOP3*	$\bar{R}^2$	*S.E.*
Argentina	− 0.93346			0.93557	0.09802
	(0.03497)				
	0.86887	− 0.26988		0.95713	0.07995
	(0.36052)	(0.05382)			
	17.58801	− 5.10160	0.45853	0.97291	0.06356
	(3.15231)	(0.90803)	(0.08610)		
Australia	− 1.96311			0.90819	0.11701
	(0.08907)				
	− 9.33423	1.16556		0.96647	0.07071
	(0.80400)	(0.12685)			
	119.80425	− 39.18070	4.16368	0.99392	0.03012
	(8.85317)	(2.76442)	(0.28523)		
Austria	− 0.87489			0.94908	0.08714
	(0.02894)				
	− 2.28214	0.24758		0.98611	0.04551
	(0.12483)	(0.02180)			
	− 5.29678	1.28843	− 0.11548	0.98844	0.04151
	(0.93756)	(0.32192)	(0.03565)		
Brazil	− 1.15342			0.98411	0.05056
	(0.01545)				
	− 2.66932	0.20821		0.99408	0.03086
	(0.12374)	(0.01695)			
	− 2.71450	0.22034	− 0.00107	0.99401	0.03103
	(1.63335)	(0.43760)	(0.03859)		
Canada	− 1.13192			0.97671	0.05893
	(0.02497)				
	1.14321	− 0.34666		0.98767	0.04287
	(0.34464)	(0.05244)			
	− 15.37158	4.61492	− 0.49283	0.99114	0.03636
	(3.76498)	(1.12859)	(0.11201)		
Colombia	− 0.84731			0.98443	0.04818
	(0.01522)				
	− 0.73050	− 0.018175		0.98428	0.04841
	(0.15824)	(0.02451)			
	3.125606	− 1.210338	0.1202176	0.98600	0.04569
	(1.48994)	(0.45890)	(0.04622)		
Czechoslavakia	− 1.10681			0.97325	0.06316
	(0.02621)				
	− 2.33225	0.20375		0.98221	0.05151
	(0.24514)	(0.04060)			
	0.34418	− 0.67639	0.094772	0.98229	0.05140
	(2.44916)	(0.80240)	(0.08629)		

TABLE 4—*Continued*

	LPOP	*LPOP2*	*LPOP3*	$\bar{R}^2$	*S.E.*
Denmark	− 1.37397			0.95373	0.08306
	(0.04321)				
	− 3.43651	0.35825		0.97111	0.06563
	(0.37888)	(0.06554)			
	19.86741	− 7.61529	0.89416	0.98646	0.04494
	(3.17455)	(1.08348)	(0.12140)		
Ethiopia	− 0.970176			0.93980	0.09475
	(0.03505)				
	− 2.65955	0.31553		0.98857	0.04128
	(0.11873)	(0.02199)			
	− 2.18810	0.13965	0.021052	0.98837	0.04164
	(1.09509)	(0.40668)	(0.04861)		
Finland	− 1.08352			0.98993	0.03876
	(0.01561)				
	− 1.23568	0.028446		0.98990	0.03881
	(0.16472)	(0.03065)			
	4.03375	− 1.90059	0.23051	0.99267	0.03306
	(1.22439)	(0.44604)	(0.05321)		
France	− 1.32463			0.95186	0.08472
	(0.04253)				
	− 4.96104	0.52137		0.99124	0.03615
	(0.24770)	(0.03542)			
	5.01212	− 2.31113	0.26489	0.99262	0.03318
	(3.19405)	(0.90543)	(0.08462)		
East Germany	− 1.12517			0.97591	0.05994
	(0.02525)				
	− 2.30689	0.18698		0.97985	0.05482
	(0.36740)	(0.05802)			
	− 22.64677	6.54614	− 0.65499	0.98815	0.04204
	(3.50383)	(1.09282)	(0.11247)		
West Germany	− 1.17093			0.98209	0.05229
	(0.02076)				
	0.78655	− 0.27644		0.98979	0.03949
	(0.29567)	(0.04170)			
	− 9.13522	2.49302	− 0.25574	0.99052	0.03804
	(4.30281)	(1.19908)	(0.11067)		
Ghana	− 1.10355			0.95453	0.08234
	(0.03439)				
	− 3.12549	0.37335		0.99067	0.03730
	(0.14869)	(0.02730)			
	− 1.91618	− 0.072435	0.053368	0.99058	0.03747
	(1.61912)	(0.59494)	(0.07115)		

TABLE 4—*Continued*

	LPOP	*LPOP*2	*LPOP*3	$\bar{R}^2$	*S.E.*
Greece	− 1.137907			0.97910	0.05583
	(0.02375)				
	− 2.488172	0.2317662		0.99129	0.03603
	(0.16418)	(0.02806)			
	4.38241	− 2.095132	0.2575483	0.99379	0.03043
	(1.54666)	(0.52225)	(0.05774)		
Hungary	− 1.092002			0.91785	0.11068
	(0.04663)				
	− 3.639834	0.4125606		0.98762	0.04296
	(0.15566)	(0.02503)			
	6.10221	− 2.709241	0.3239994	0.99346	0.03124
	(1.49145)	(0.47690)	(0.04946)		
India	− 1.20419			0.99498	0.02905
	(0.00703)				
	− 1.54161	0.04652		0.99535	0.02796
	(0.09497)	(0.01306)			
	− 1.52389	0.04172	0.00043	0.99532	0.02806
	(1.20827)	(0.32641)	(0.02908)		
Indonesia	− 0.96680			0.99021	0.03820
	(0.01373)				
	− 1.65593	0.09780		0.99324	0.03175
	(0.14575)	(0.02062)			
	− 1.66667	0.10079	− 0.00027	0.99309	0.03209
	(1.62185)	(0.45020)	(0.04111)		
Iran	− 0.9926936			0.97808	0.05717
	(0.02122)				
	− 1.916886	0.1436970		0.98599	0.04570
	(0.17515)	(0.02711)			
	4.824663	− 1.891429	0.2004492	0.98979	0.03901
	(1.57453)	(0.47373)	(0.04660)		
Israel	− 0.9829362			0.98682	0.04433
	(0.01622)				
	− 0.3794996	− 0.1133325		0.98946	0.03964
	(0.16791)	(0.03142)			
	− 1.937064	0.4663606	− 0.07054118	0.98940	0.03976
	(1.83547)	(0.68097)	(0.08278)		
Italy	− 1.046321			0.98452	0.04804
	(0.01874)				
	− 1.706679	0.09282604		0.98550	0.04649
	(0.32065)	(0.04500)			
	− 28.13986	7.459235	− 0.6776479	0.99636	0.02330
	(2.23031)	(0.62034)	(0.05703)		

TABLE 4—*Continued*

	LPOP	*LPOP2*	*LPOP3*	$\bar{R}^2$	*S.E.*
Japan	− 1.28916			0.97920	0.05898
	(0.01605)				
	− 2.33129	0.14177		0.98324	0.05294
	(0.17977)	(0.02438)			
	9.07647	− 2.89048	0.26499	0.98779	0.04518
	(1.59937)	(0.42367)	(0.03698)		
Malaysia	− 0.9678605			0.96279	0.07449
	(0.02717)				
	0.6007225	− 0.2920827		0.98383	0.04910
	(0.19774)	(0.03667)			
	6.950913	− 2.621211	0.2793366	0.98651	0.04485
	(1.98319)	(0.72515)	(0.08688)		
Mexico	− 1.15342			0.98520	0.04697
	(0.02019)				
	− 2.23244	0.15499		0.98982	0.03895
	(0.22661)	(0.03246)			
	8.78962	− 2.97678	0.29295	0.99297	0.03238
	(2.35625)	(0.66790)	(0.06243)		
Morocco	− 0.8090991			0.98201	0.05180
	(0.01564)				
	− 0.5191530	− 0.04982536		0.98284	0.05058
	(0.15958)	(0.02730)			
	− 1.423535	0.2520110	− 0.0327586	0.98267	0.05084
	(1.25371)	(0.41589)	(0.04504)		
Netherlands	− 1.266138			0.97600	0.05983
	(0.02836)				
	− 2.411108	0.1802397		0.97914	0.05577
	(0.39979)	(0.06280)			
	12.72685	− 4.603166	0.4985219	0.98118	0.05297
	(6.14550)	(1.93911)	(0.20200)		
Nigeria	− 1.536646			0.97384	0.06245
	(0.03597)				
	− 0.7895532	− 0.1167853		0.97402	0.06225
	(0.65101)	(0.10161)			
	51.96309	− 16.43155	1.670384	0.98268	0.05083
	(10.67219)	(3.29753)	(0.33751)		
Norway	− 1.265160			0.96694	0.07021
	(0.03341)				
	− 2.927434	0.3060476		0.98257	0.05098
	(0.25158)	(0.04610)			
	7.88985	− 3.60444	0.46173	0.98827	0.04183
	(2.22617)	(0.80219)	(0.09461)		

TABLE 4—*Continued*

	LPOP	*LPOP2*	*LPOP3*	$\bar{R}^2$	*S.E.*
Philippines	− 1.25315			0.97092	0.06586
	(0.03098)				
	− 1.96262	0.109152		0.97205	0.06455
	(0.41390)	(0.06351)			
	17.02205	− 5.72756	0.59109	0.98133	0.05276
	(3.86179)	(1.18386)	(0.11977)		
Poland	− 1.12680			0.98689	0.04421
	(0.01855)				
	0.54508	− 0.25506		0.99463	0.02830
	(0.20000)	(0.03046)			
	− 3.88054	1.077172	− 0.13243	0.99483	0.02776
	(2.62377)	(0.78818)	(0.07829)		
Romania	− 1.08548			0.93326	0.09976
	(0.04144)				
	− 1.61719	0.08707		0.93392	0.09927
	(0.43977)	(0.07170)			
	14.03749	− 4.83611	0.50608	0.94380	0.09154
	(5.15842)	(1.61858)	(0.16624)		
South Africa	− 0.99689			0.95940	0.07781
	(0.02928)				
	− 0.22443	− 0.12402		0.96109	0.07617
	(0.44047)	(0.07057)			
	− 11.06859	− 3.39352	− 0.37583	0.96197	0.07530
	(7.51194)	(2.43356)	(0.25990)		
Spain	− 1.13272			0.97349	0.06287
	(0.02669)				
	− 2.70278	0.22313		0.98427	0.04843
	(0.27053)	(0.03834)			
	9.20500	− 3.14134	0.31272	0.98757	0.04305
	(3.25078)	(0.91661)	(0.08514)		
Sri Lanka	− 1.13030			0.98018	0.05437
	(0.02296)				
	− 2.15554	0.18901		0.98751	0.04316
	(0.19065)	(0.03499)			
	1.032618	− 0.95712	0.13442	0.98808	0.04217
	(1.78519)	(0.63918)	(0.07486)		
Sweden	− 1.40997			0.95200	0.08460
	(0.04520)				
	− 5.18718	0.61077		0.98357	0.04950
	(0.39218)	(0.06327)			
	− 9.99120	2.15972	− 0.16461	0.98343	0.04971
	(6.13961)	(1.97652)	(0.20994)		

TABLE 4—*Continued*

	LPOP	*LPOP2*	*LPOP3*	$\bar{R}^2$	*S.E.*
Switzerland	− 1.09535			0.96523	0.07201
	(0.02969)				
	− 2.246574	0.20911		0.96961	0.06732
	(0.41006)	(0.07431)			
	− 24.03732	8.034563	− 0.92286	0.98750	0.04317
	(2.64955)	(0.94800)	(0.11166)		
Thailand	− 0.96098			0.83206	0.15825
	(0.06155)				
	− 4.28744	0.54068		0.96034	0.07690
	(0.26779)	(0.04325)			
	− 5.90579	1.08853	− 0.05972	0.95982	0.07741
	(2.61529)	(0.88169)	(0.09600)		
Turkey	− 1.07716			0.97323	0.06319
	(0.02552)				
	− 2.44627	0.20588		0.98440	0.04823
	(0.23097)	(0.03461)			
	6.56527	− 2.49188	0.26487	0.98728	0.04356
	(2.65130)	(0.79187)	(0.07769)		
U.K.	− 1.17828			0.94880	0.08832
	(0.03624)				
	− 3.44720	0.30838		0.97868	0.05699
	(0.25558)	(0.03459)			
	21.88441	− 6.41464	0.58460	0.99605	0.02452
	(1.62861)	(0.43151)	(0.03750)		
USSR	− 1.27808			0.97607	0.06436
	(0.01337)				
	0.31357	− 0.22347		0.98292	0.05437
	(0.16773)	(0.02350)			
	13.14734	− 3.68450	0.30757	0.98711	0.04724
	(1.50855)	(0.40544)	(0.03598)		
USA	− 1.18387			0.98623	0.04816
	(0.01142)				
	− 0.77942	− 0.05602		0.98680	0.04715
	(0.14851)	(0.02051)			
	9.59841	− 2.82653	0.24321	0.98982	0.04141
	(1.55574)	(0.41426)	(0.03633)		
Venezuela	− 1.10561			0.99204	0.03446
	(0.01415)				
	− 0.43973	− 0.10722		0.99395	0.03003
	(0.16595)	(0.02665)			
	6.71913	− 2.38959	0.23942	0.99561	0.02558
	(1.65869)	(0.52739)	(0.05527)		

TABLE 4—*Continued*

	LPOP	*LPOP*2	*LPOP*3	$\bar{R}^2$	*S.E.*
Yugoslavia	− 1.18561			0.98544	0.04660
	(0.02058)				
	− 2.43190	0.20449		0.99017	0.03828
	(0.25427)	(0.04163)			
	− 14.51490	4.13860	− 0.42184	0.99270	0.03300
	(2.92005)	(0.94874)	(0.10166)		
Zaire	− 0.92952			0.99379	0.03043
	(0.01050)				
	− 0.87006	− 0.01008		0.99370	0.03065
	(0.10885)	(0.01836)			
	2.99662	− 1.29615	0.13946	0.99579	0.02505
	(0.78869)	(0.26108)	(0.02826)		

Ijiri and Simon [5] have argued that the curvature of rank-size graphs can arise without a violation of Gibrat's law, suggesting that the curvature may result from autocorrelation of growth rates over time. Ijiri and Simon also argue that mergers and aquisitions can cause a curvature in rank-size plots of firms. This analysis, applied to city sizes, would mean that because of suburbanization, the largest *cities* might be growing more slowly and so downward curvature might be expected. Vining [9], contended that it was not autocorrelated growth itself, but rather other features of the Ijiri–Simon model—specifically, the inverse relationship between growth rates and age—which were responsible for observed curvatures. To reach this conclusion, however, Vining used the accepted empirical fact that rank-size graphs have a downward concavity. He demonstrated that autocorrelated growth by itself (in the absence of any relation between growth and age) causes an upward concavity. He concluded that growth rates must be negatively related to size or age in order to "offset" this upward concavity.

Vining's argument loses validity if most rank-size graphs have upward concavity, as we have found. Autocorrelated growth does not introduce the "wrong" concavity, as Vining argued. In fact, it seems reasonable that autocorrelated growth and a positive relation of growth rates to size are the primary factors combining to give upward concavity to most rank-size graphs. For only about one-third of the countries we studied do diseconomies of scale and age offset or dominate these factors to give downward concavity to the graphs.

Despite these deviations from linearity, the Pareto distribution may be the best general description of rank-size data. Several other distributions, notably the lognormal, have been suggested as alternatives to the Pareto. The upward concavity we observed, however, serves as evidence against the lognormal distribution, since such a distribution is characterized by downward concavity, especially toward the lower end (smaller populations) of the distribution (Parr and Suzuki [7]). With only 50 cities included in most of our samples, it is possible that we would not observe extreme downward concavity even if the overall distribution were lognormal, since the upper tail of the lognormal is almost linear. However, if the overall distribution were indeed lognormal, we would nonetheless expect to find the upper tails of our samples tending to be concave downward in more cases than concave upward.

One possible functional form which remains is the S-shaped curve proposed by Stewart [8]. The major portion of this curve is concave downward, but the largest cities are disproportionately large, causing the upper tail to curve concave upward. Our 50-city samples may have revealed only the upper tail, masking the downward concavity of the rest of this distribution.

5. FACTORS EXPLAINING VARIATIONS IN PARETO EXPONENT AND PRIMACY

Pareto exponents and measures of primacy varied widely among the 44 countries studied. In an attempt to explain this variation, a cross-section analysis was done. We excluded Australia from our sample for these tests, considering it to be an exceptional case.

A number of alternative variables and specifications were utilized. Four key variables seemed to be consistently significant: total population, area, GNP per capita, and railway mileage density (railway mileage divided by area). The two preferred equations follow:

$$\lambda = 0.9800 + \underset{(0.01783)}{0.05931\, GNPCAP} + \underset{(0.0644)}{0.1460\, POP^{1/2}} - \underset{(0.0500)}{0.1017\, AREA} + \underset{(0.0059)}{0.1134\, (AREA)^2},$$

$$\bar{R}^2 = 0.2197 \qquad \text{Standard error} = 0.1358 \tag{4}$$

$$\lambda = 1.0054 + \underset{(0.0269)}{0.0832\, GNPCAP} + \underset{(0.0700)}{0.1794\, (POP)^{1/2}} - \underset{(0.0633)}{0.1479\, AREA} + \underset{(0.0068)}{0.0155\, (AREA)^2} - \underset{(0.2353)}{0.2779\, (RAIL)^{1/2}},$$

$$\bar{R}^2 = 0.2281, \qquad \text{Standard error} = 0.1350. \tag{5}$$

In interpreting these results, we associate higher Pareto exponents with more evenly distributed populations. The GNP per capita and population coefficients indicate that wealthier and more populous countries have more evenly distributed populations. In heavily populated countries, intermediate-sized cities are likely to develop since the major cities may reach a size where negative externalities discourage growth in the cities. High income allows a country to support a network of intermediate-sized cities as demand thresholds are passed. In both cases, the overall population becomes more evenly distributed and the Pareto exponent rises.

The "curve" indicated by the two measures of geographic area in the regression decreases through most of the points in the sample, then turns up in the range of very large countries. This primarily negative correlation is somewhat counter-intuitive; we would expect countries with larger areas to be more evenly populated. This is, however, true for very large countries like the U.S.S.R. and the U.S.A.

In (5) the density of the rail-network was added but was not statistically significant. We also tried adding a variety of other variables to the regression equation—exports/GNP, manufactured exports/total exports, growth rate, nonagricultural labor force/total labor force, and dummy variables for non-market economies and for former colonies—but none of these variables entered significantly.

We also tested to see if our sample might be better understood as two "subsamples." Previous theories have described the rank-size rules as the result of an equilibrium process, whereby the Pareto exponent drifts toward unity (from either above or below) as a country develops. We used the mean Pareto exponent of 1.117 (excluding Australia) as a dividing point to create two subsamples, and we ran regressions for each of the two subsamples. Our results proved insignificant. We then ran a regression using the absolute difference between a country's Pareto exponent and the mean of 1.117 as dependent variable. The best equation is

$$|\lambda - 1.117| = 0.1900 - \underset{(0.1341)}{0.3391}\,(RAIL)^{1/2} + \underset{(0.00375)}{0.00613}\,(GNPCAP)^2 - \underset{(0.0335)}{0.0653}\,AREA + \underset{(0.0041)}{0.0071}\,(AREA)^2$$

$$\bar{R}^2 = 0.0682, \qquad \text{Standard error} = 0.09656. \tag{6}$$

These results are comparable to our earlier regressions.

As a comparison to these Pareto exponent results, we ran similar regressions using two measures of primacy as dependent variables. Using population, area, GNP per capita, and rail density as independent variables, the best equations follow.

$$PRIME1 = 0.5748 - \underset{(0.0433)}{0.0565}\,(POP)^{1/2} - \underset{(0.0351)}{0.0657}\,(GNPCAP)^{1/2} + \underset{(0.9073)}{1.7289}\,(RAIL)^2$$

$\bar{R}^2 = 0.07845$, Standard error $= 0.11359$.

$$PRIME2 = 0.3413 - \underset{(0.0344)}{0.0578}\,(POP)^{1/2} - \underset{(0.0279)}{0.0734}\,(GNPCAP)^{1/2} + \underset{(0.7206)}{1.2911}\,(RAIL)^2$$

$\bar{R}^2 = 0.13915$, Standard error $= 0.09020$.

The signs of the population and GNP terms in these equations are as expected. In wealthier and more populated countries, one would expect that a single primate city could not dominate. The railroad term indicates that the countries with a better rail system are more likely to be primate. This is reasonable since an efficient transportation system enables one city to service a larger hinterland. Area did not enter significantly in either equation.

An additional variable was included in the primacy equations to test the hypothesis proposed by a number of economists concerning the relationship of export dependency and primacy. The percentage of GNP originating in the export sector was included in both equations.* Our results indicate that greater export dependency reduces primacy, results which contradict the conventional wisdom. Our equations follow.

$$PRIME1 = 0.7301 - \underset{(0.0493)}{0.1084}\,(POP)^{1/2} - \underset{(0.0339)}{0.0684}\,(GNPCAP)^{1/2} + \underset{(0.8770)}{1.8576}\,(RAIL)^2 - \underset{(0.05285)}{0.1045}\,(EXPORT)^{1/2},$$

$\bar{R}^2 = 0.14395$, Standard error $= 0.10948$.

$$PRIME = 0.4040 - \underset{(0.0375)}{0.0884}\,(POP)^{1/2} - \underset{(0.0272)}{0.0760}\,(GNPCAP)^{1/2} + \underset{(0.7007)}{1.3284}\,(RAIL)^2 - \underset{(0.0153)}{0.0275}\,EXPORT$$

$\bar{R}^2 = 0.18676$, Standard error $= 0.08767$.

6. CONCLUSION

This paper has examined the Pareto and primacy measures of the size distribution of cities. The mean Pareto exponent for a sample of 44 countries is 1.136, somewhat greater than the exponent of one implied by the rank-size rule. However, we find that value of the Pareto exponent is quite sensitive to the definition of the city and the choice of sample size. The significance of non-linear terms in variants of the Pareto distribution also indicate that the rank-size rule is only a first approximation to a complete characterization of the size distribution of cities within a country. The relatively low correlation between primacy and Pareto measures confirm the need for a variety of measures of city size distributions.

This paper also suggests that large cities are growing faster than small cities in most of the countries in our sample. This is indicated by the positive coefficient on the first non-linear term introduced into the Pareto equation.

Finally, variations in the Pareto exponent and measures of primacy are partly explained by economic, demographic, and geographic factors. The empirical work in this and other studies lacks one crucial element, a rigorous theoretical model explaining the size distribution of cities. Hopefully, this empirical exploration will encourage such work.

APPENDIX: Sources of Data

The city population data for this study was obtained from four sources:

(1) United Nations Demographic Yearbooks,

(2) Rand McNally Commercial Atlas and Marketing Guide,

(3) Information from Hammond Inc., a company which publishes atlases,

(4) Statistical abstracts for individual countries.

For countries with at least 50 cities larger than 50,000, we relied primarily on the first two sources. For less-populous countries, it was necessary to use the latter two sources. We attempted to find urban population figures from as close to 1970 as possible. For 43 of the 44 countries studied, we were able to find data (either census figures or official estimates) representing a year between 1966 and 1973. (Nigeria, for which we used a 1963 census, was the only exception.)

Following is a list of the years and sources of the city population data for each of the 44 countries:

Argentina	1970	Rand McNally
Australia	1970	Rand McNally
Austria	1971	statistical abstract

Brazil	1970	Rand McNally
Canada	1971	Rand Mcnally
Colombia	1973	Hammond
Czechoslavakia	1970	Hammond
Denmark	1971	statistical abstract
Ethiopia	1971	Hammond
Finland	1970	statistical abstract
France	1970	Rand McNally
East Germany	1970	statistical abstract
West Germany	1970	UN Demographic Yearbook
Ghana	1970	statistical abstract
Greece	1971	statistical abstract
Hungary	1970	statistical abstract
India	1971	UN Demographic Yearbook
Indonesia	1971	Rand McNally
Iran	1966	Hammond
Israel	1972	statistical abstract
Italy	1971	Rand McNally
Japan	1970	Rand McNally
Malaysia	1970	Hammond
Mexico	1970	Rand McNally
Morocco	1971	Hammond
Netherlands	1970	statistical abstract
Nigeria	1963	Rand McNally
Norway	1970	statistical abstract
Philippines	1970	Hammond
Poland	1970	Hammond
Romania	1970	statistical abstract
South Africa	1970	Hammond
Spain	1970	Rand McNally
Sri Lanka	1971	Hammond
Sweden	1970	statistical abstract
Switzerland	1970	statistical abstract
Thailand	1970	statistical abstract
Turkey	1970	statistical abstract
U.K.	1971	Rand McNally
USSR	1970	UN Demographic Yearbook
USA	1970	UN Demographic Yearbook
Venezuela	1971	Hammond
Yugoslavia	1971	Hammond
Zaire	1970	Hammond

For the cross-sectional analysis, all of the data were obtained from United Nations Statistical and Demographic Yearbooks, except for the GNP/capita figure (which came from the World Bank Atlas) and area figures (which came from the Associated Press Almanac).

REFERENCES

1 M. J. Beckmann, City hierarchies and the distribution of city size, *Econ. Develop. and Cultural Change*, **6**, No. 3, 243–248 (1958).

2 B. J. L. Berry, Contemporary urbanization processes, *in* "Geographical Prospectives and Urban Problems" (Frank E. Horton, Ed.), (a symposium organized by the Committee on Geography of the Division of Earth Sciences, National Academy of Sciences, 1971), National Academy of Sciences, Washington, D.C. (1973).

3 B. J. L. Berry, Cities as systems within systems of cities, *Papers Region. Sci. Assoc.*, **13**, 147–163 (1964).

4 B. J. L. Berry, City size distribution and economic development, *Econ. Develop. and Cultural Change*, **9**, No. 4, 573–587 (1961).

5 Y. Ijiri and Herbert A. Simon, Interpretation of departures from the Pareto curve firm-size distributions, *J. Pol. Econ.*, March–April (1974).

6 E. S. Mills, "Urban Economics," Scott, Foresman, Glenview, Ill. (1972).

7 J. Parr and K. Suzuki, Settlement populations and the lognormal distribution, *Urban Studies* **10**, 335–52 (October 1973).

8 C. T. Stewart, The size and spacing of cities, *Geographical Rev.*, **48**, No. 2, 222–245 (1958).

9 D. Vining, Autocorrelated growth rates and the Pareto law: A further analysis, *J. Pol. Econ.*, April (1976).