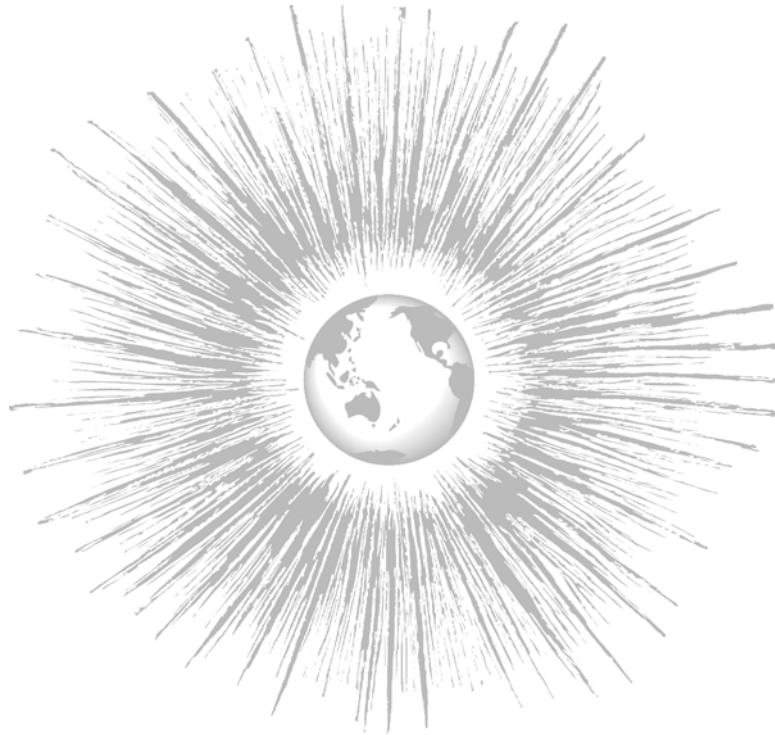


THE CITY-COUNTRY RULE: AN EXTENSION OF THE RANK-SIZE RULE

Rein Taagepera
Edgar Kaskla



ABSTRACT

A “city-country rule” derived from the well-known rank-size rule for cities correctly predicts the average relationship between the population of a country’s R th-ranking city (P_R) and the total population (P) of that country:

$$P = R P_R \ln(R P_R).$$

The formula applies in principle to other systems of defined total size, such as revenues of firms within industries or citations of journals within a scholarly discipline. The important result is that we can thus predict the populations of all cities in a country, once its total

population is given, and quickly tell whether the actual city populations are large or small, compared to the world average for similarly ranked cities in countries of comparable size. For countries of more than 100 million, a more elaborate formula holds. The largest city in the country needs a correction (primacy factor), which vanishes as the country population increases. Actual primacy may be relative (compared to other cities) or absolute (compared to country population).

INTRODUCTION

This study introduces a “city-country rule” to complement the well-known rank-size rule for cities, from which it is derived. The city-country rule enables us to make a rough estimate of the population of the largest cities when the population of the entire country is known. It quickly tells us whether the actual city populations are large or small, compared to the world average for similarly ranked cities in countries of comparable size.

The existing rank-size rule describes the empirical relationship between a city’s or town’s population and its ranking relative to other cities within an interacting geographical area. This regularity was first noted by Auerbach (1913) and later popularized by Stewart (1947) and Zipf (1949). Hence it is often called Zipf’s law. If R is the rank in size of a given city and P_R is its population, the simplest form of the rank-size rule is

$$R P_R = \text{constant} = P_1(1)$$

where P_1 is the population of the largest city. When graphed on doubly logarithmic paper, P_R versus R , this equation corresponds to a straight line with slope -1 . For many countries (or other interacting regions) a different slope ($-n$) yields a better data fit, and the corresponding generalized expression is

$$R^n P_R = \text{constant} = P_1^n(2)$$

As an empirical observation, the rule extends far beyond city sizes. It has been claimed to apply to the frequency of English words, the frequency of cita-

Rein Taagepera
Department of Political Science
University of California
Irvine, CA 92697
rtagepe@uci.edu
<http://hypatia.ss.uci.edu/ps/>

Edgar Kaskla
Department of Political Science
California State University, Long Beach
EKaskla@aol.com
<http://front.csulb.edu/politicalscience/>

JOURNAL OF WORLD-SYSTEMS RESEARCH, VII, 2, FALL 2001, 157–173

<http://jwsr.ucr.edu>

ISSN 1076-156X

© 2001 Rein Taagepera & Edgar Kaskla

tion of scholarly journals, the size distribution of organisms in an ecosystem, the revenues of firms in an industry, and even the magnitude of earthquakes. We prefer to call it a “rule” rather than “law,” because no widely accepted theoretical foundation has been supplied, thus it remains an empirical rule rather than a scientific law.

An extensive literature has attempted to explain the existence of such a rank-size relationship. Berry (1966) in particular has drawn upon the earlier work of Simon (1955) to suggest that the rank-size regularity is based on purely mathematical grounds through stochastic growth models. But other works (Vining 1974, Okabe 1977) have challenged the mathematical assumptions that produce such a “steady state” model. Further approaches (Dacey 1966, Beckmann and McPherson 1970, Mulligan 1976, Beguin 1979) have focused on the assumptions that produce a continuous rank-size relationship rather than a stepped or hierarchical relationship that would be consistent with central place theory as developed by Christaller (1966).¹

Empirical testing has also continued on the level of individual countries and regions past and present (Berry and Kasarda 1977, Skinner 1977, Alden 1979, Asami 1986, Vining 1986) as well as for the world as a whole (Chase-Dunn 1985, Ettlinger and Archer 1987). Agreements and disagreements with the rank-size rule have been observed, as well as difficulties in testing it, since “legal” (administrative) and actual sizes of population centers differ (Asami 1986) and national units may not always be the most suitable units in which the test should be carried out.

Further operational difficulties have been noted in defining (1) the limits of a city system and (2) the precise meaning of “primacy” (Walters 1985). The latter term expresses the observation that the rule at times works with all other cities except the largest one, which tends to be larger than expected. This “primacy factor” will be discussed in more below.

Now we come to the new idea investigated in this study. The rank-size rule has been a rule about sizes of components (such as cities). But whenever these components add up to a definable total, it also has implications regarding this

¹ There are only three relatively simple patterns to choose from, when one wants to have a steady decrease in size: linear, power rule (Eq. 2) and exponential. Any actual distribution of components ranked by size is bound to approach one of these three, more or less perfectly. We observe, for instance, that the seat distribution of parties in representative assemblies tends to follow an exponential pattern rather than a power rule (Taagepera 2001). The broad theory to explain these differences remains to be constructed.

total. Cities and other smaller settlements do add up to the total population of the country (or other territorial unit).² Our purpose here is to make this implicit relationship explicit and to use it in predicting population sizes of cities in a given state system. By so doing, we will expand the scope of the rank-size rule considerably. This rule is all too often dismissed as being an interesting empirical regularity without much applicability as a predictive tool. The new “city-country rule” does predict the sizes of all cities in an average system, on the basis of total population. Deviations from the predictions indicate that the actual system is different from this average. Primacy, too, can be defined and measured in a more precise way when the total population of the country is taken as the base line.

THE MODEL FOR THE CITY-COUNTRY RULE

Assume that the rank-size rule (Eq. 2) applies to all settlements, down to the last isolated one-person dwelling.³ By summing up all these populations, one should get the population of the entire country. The details are worked out in Appendix A. If $n=1$ (Eq. 1), the following equation expresses the population P of the entire country (or other suitable territorial unit) in terms of the population P_R of its R -th largest city or settlement:

$$P = R P_R \ln(R P_R). \quad (3)$$

For the largest city in the country ($R=1$), this becomes

$$P = P_1 \ln P_1 \quad (4)$$

and for the next-largest city ($R=2$), it is

$$P = 2 P_2 \ln(2 P_2). \quad (5)$$

The general equation (corresponding to Eq. 2, with any slope indicator n) is more complex:

² The same is the case for revenues of firms in an industry and many other phenomena. It is not the case, however, for others, such as magnitude of earthquakes.

³ Here another assumption enters: that a conceivable smallest possible size can be defined. This is the case for human settlements (one person), for words in a defined set (one single occurrence), and for the journals in a set of citations (one single occurrence). This is not the case for revenues of firms, where the cutoff between zero and some minimal revenue is arbitrary, depending on the currency used. Thus the city-country rule is based on two assumptions: that the total of the components can be defined, and that the components themselves consist of discrete particles of one type (persons, in the case of cities).

$$P = R P_R (P_R^{-1+1/n} - R^{n-1}) / (1-n). \quad (6)$$

In the case of the largest city ($R=1$), this yields

$$P = P_1 (P_1^{-1+1/n} - 1) / (1-n), \quad (7)$$

and for the next-largest city ($R=2$), it becomes

$$P = 2P_2 (P_2^{-1+1/n} - 2^{n-1}) / (1-n). \quad (8)$$

We will test the city country rule with worldwide data for the largest and second-largest cities (Equations 4, 5, 7 and 8), as well as for the third-, fourth- and fifth-ranking cities (using the analogous equations derived from Eqs. 3 and 6). The population of the largest city often deviates from the rank-size rule, which sets in with the next-largest city. Therefore, the relation between the populations of a country and its second-largest city (and the third-, fourth- and fifth-largest) is expected to give a better fit than is the case with the largest city. We will start the test with these next-largest cities. Only if this first test turns out fairly successful would there be any point in testing for the largest cities, so as to evaluate the extent of the primacy factor.

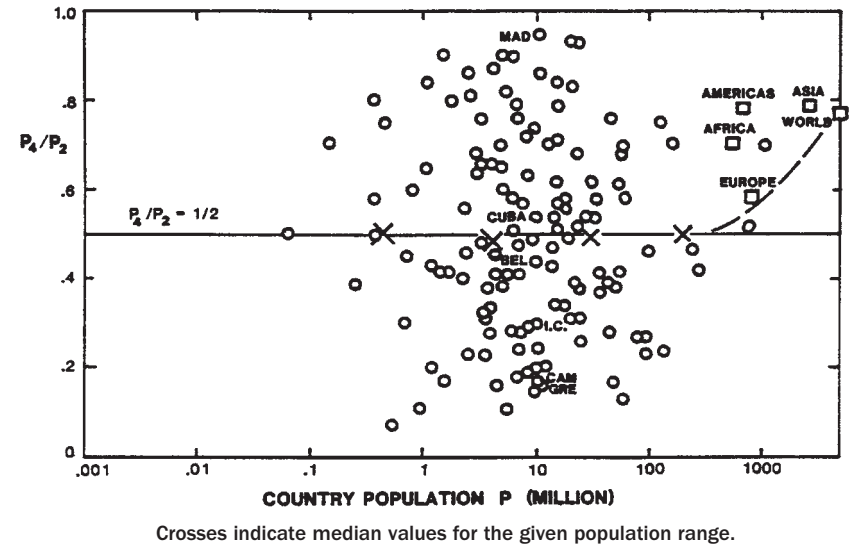
TESTING THE CITY-COUNTRY RULE

In order to test the implications of the rank-size rule for the total population of countries, we collected data for 203 formally independent or effectively autonomous countries as given in *The World in Figures* (1988). The data refer to the period around 1985 and include the populations of countries and up to five largest cities, with increasing gaps for the lower ranking cities, especially for countries of less than half a million people. For the fifth-largest cities the number of cases decreases to 124.

Before proceeding to the direct test, we should determine the median value of power index n (assuming that countries do follow the rank-size rule). This is done in the following way. Figure 1 shows the population ratio P_4/P_2 of the fourth- and second-largest cities graphed against the population of the entire country.⁴ According to Eq. 2, the median ratio should be $1/2^n$. While the points are extremely scattered, the median values of P_4/P_2 (indicated as x-es in Figure 1) are strikingly close to $1/2$, at any country population below 300 million. This suggests that, on the average, the same value of n applies to countries of almost

⁴ Why pick the fourth- and second-largest cities? We want to avoid the largest cities, where the primacy factor enters. Beyond the fourth-largest cities, gaps in our data set develop. Within this range we want to have sufficient contrast; hence P_4/P_2 is chosen. Countries at exactly 10 million—Belgium, Cameroon, Cuba, Greece, Ivory Coast and Madagascar—are labeled in Figure 1, because we will use them in a later section.

Figure 1 – Population ratio of fourth- and second-largest cities as a function of the total population of the country.



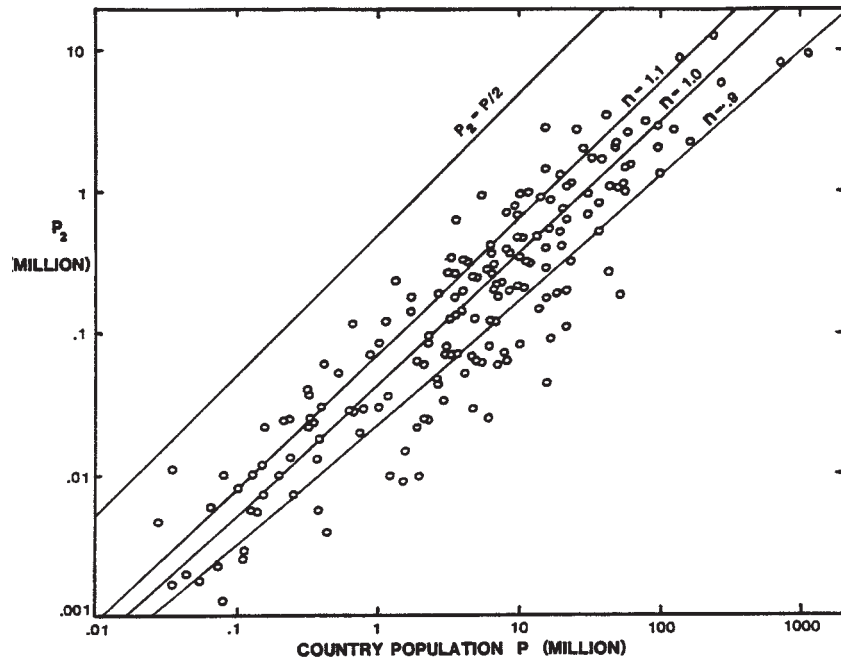
any size, and this value is $n=1.0$, in line with the simplest form of the rank-size rule (Eq. 1). For the world as a whole (as also shown in Figure 1) a lower value of $n=0.85$ would apply, and the same tends to be the case for continents.

In this light, what could be considered a successful test of the city-country model? If we graphed the population P_2 of the second-largest city versus the country population P , all points should fall reasonably close to the curve corresponding to $n=1.0$ (as given by Eq. 5). In order to be consistent with the findings of Figure 1, we would wish for most data points to fall in the narrow zone delineated by the curves for $n=0.9$ and $n=1.1$ (as determined by using Eq. 8). Moreover, this should be especially the case for country populations below 300 million. For the larger countries, a deviation toward lower values of n is expected in the light of Figure 1.

Figure 2 shows P_2 (the size of the second-ranking city) graphed against P (the population of the country), both on logarithmic scales. The theoretical curves for $n=0.9$, 1.0 and 1.1 (based on Eq. 5 and 8) are shown.⁵ The actual data

⁵ These curves are not straight lines on the log-log graph, but the deviation from straight lines is almost invisible throughout the range shown. For reference, the line $P_2=P/2$ is also shown. This is the maximum conceivable size the second-ranking city could have; it would be reached when a country consists of just two cities with equal populations.

Figure 2 – Population of the second-largest city as a function of the total population of the country.



points are visibly most crowded in the zone between the $n=0.9$ and $n=1.1$ curves, and $n=1.0$ is close to the best fit curve for these data. Countries with very large populations (India and China) are on the low side, as expected. Thus our simple model does agree, indeed, with the pattern observed.

A similar test was carried out for the third-, fourth- and fifth-largest cities. The patterns of the corresponding graphs (not shown here) are quite similar to that in Figure 2 in that most country points fall between $n=0.9$ and $n=1.1$. As shown at the bottom of Table 1, more than 50% of the data points for the second-largest cities are located within the zone between $n=0.9$ and $n=1.1$. For the next-largest cities, this proportion gradually increases to 64%.

The main body of Table 1 lists the median values of power constants (n) at various ranks (from $R=1$ to $R=5$). This is done separately in various country population brackets. The median n for all cities in all countries is 0.99. The first-ranking cities will be discussed later. For all lower-ranking cities the values of n are within plus or minus 0.07 of $n=1.00$, as long as the total populations of countries remain below 100 million. For countries larger than 100 million, the

Table 1 – Median Power Constants at Various Total Populations of Countries for Largest to Fifth-largest Cities

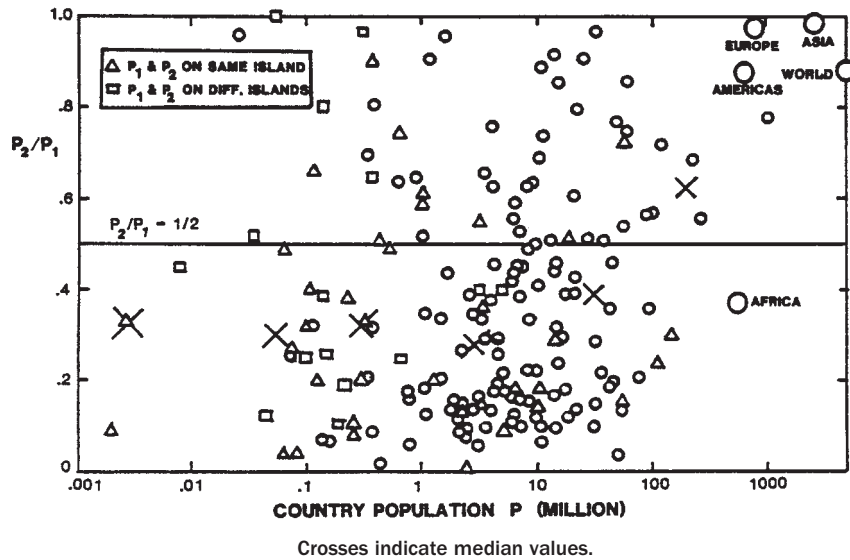
Country population (million)	$n(P_1)$	$n(P_2)$	$n(P_3)$	$n(P_4)$	$n(P_5)$
.01 to 1	1.26	.96	—	—	—
.1 to 1	1.19	1.07	1.01	1.00	1.00
1 to 10	1.08	.98	.94	.96	.98
10 to 100	1.01	.97	.99	1.00	.98
Over 100	.98	.96	.93	.92	.94
Median of All Countries	1.08	.98	.97	.97	.96
The World	.81	.84	.84	.86	.87
Percentage of Countries where $0.9 < n < 1.1$	49	51	54	60	64

values of n are consistently below 1.0, with a mean of 0.94, and for the world as a whole the values of n drop to around 0.85. The median n for all second-ranking cities is 0.98, and for the next-ranking cities it slowly but consistently decreases to 0.96 for the fifth-ranking cities.

Thus, the average relationship between the populations of a country and of its second- to fifth-largest cities is well predicted by our model, despite its very simple starting point. Except for very large countries, the value of the parameter n that yields the best overall fit (.99) is extremely close to 1.00.

We may well wonder whether the simple relationship with $n=1$ is some theoretical norm for city systems which are genuinely separate interacting systems rather than parts of a larger system or composites of several systems. In this light, the lower n for very large countries could mean that they consist of several separate geographical systems. If so, then n should be especially low if the entire world were tentatively considered a single system. This is, indeed, the case (see Table 1): for the world, n is around 0.85. Going in the reverse direction, the tiniest countries still do not have n appreciably larger than 1.0, suggesting that, on the average, they are separate systems rather than parts of larger regional systems.

Figure 3 – Population ratio of second-largest and largest cities as a function of the total population of the country.



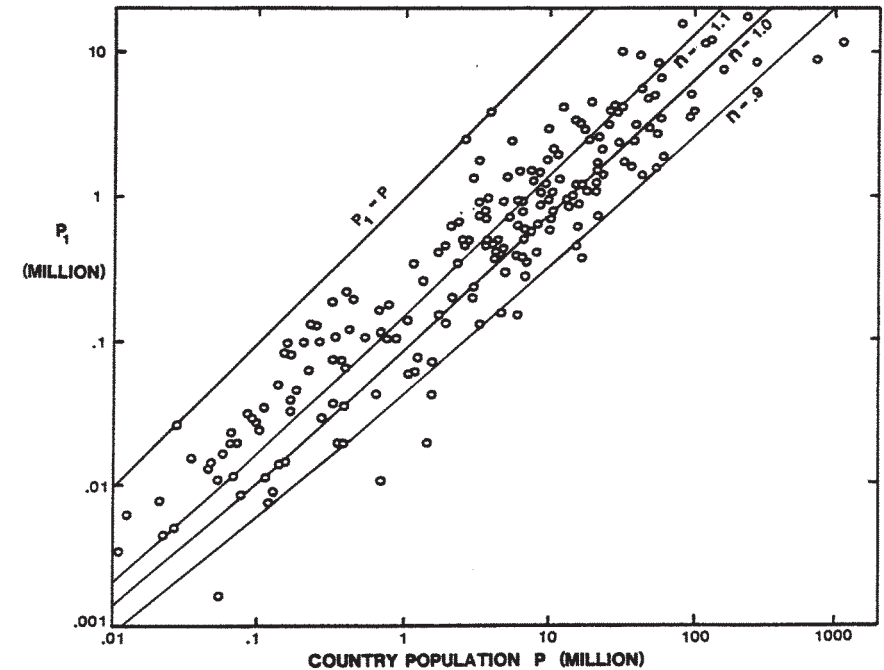
THE PRIMACY FACTOR

It is now time to apply the model to the first-ranking cities. Because of the primacy factor, we expect a shift away from the $n=1.0$ curve. As a test of these expectations, we will again use the format of Figure 1. Figure 3 shows the population ratio P_2/P_1 of the largest and the second-largest cities graphed against the population of the entire country. (The island countries, marked with a special symbol, are discussed in Note 6.) If there were no primacy effect, we should have the same median n as in Figure 1 ($n=1.0$), and hence a median $P_2/P_1=0.5$. This is visibly not the case. The median values of P_2/P_1 at various P (indicated as x -es in Figure 3) are clearly below the 0.5 level, except for country populations above 100 million.

What this means is that for country populations of less than 100 million there is a primacy effect. It tends to be most marked when country population remains below 10 million. For country populations of more than 100 million the primacy effect vanishes or even reverses itself, in a pattern similar to that observed in Figure 1.

The primacy question has often arisen, also in relation to theories of under-development and dependency (see Ettliger and Archer 1985). Our test suggests

Figure 4 – Population of the largest city as a function of the total population of the country.



that primacy is largely the consequence of being a small country. Within the present context of transnational economic interdependence (whatever its causes) and communication linkages, the largest city in a small country must be in close contact with other large cities in a wider region of the world. These strong connections to centers of other countries present economic opportunities that provide their own impetus for further development of the largest city in a small country, leading to a hierarchy of central places that follows the pattern suggested by Christaller (1966) and Lösch (1954). All other things being equal, the development of central places must be analyzed in terms of the location vis-a-vis other central places, and so too the ability of a threshold population to support the variety of activities found in the highest-ranked centers. Political borders play an important role in small states, because they artificially limit the area to be serviced. Within that area the capital city may turn out to have a greater variety of services than the major city of an equally large area within a larger country.⁶

⁶ – and archipelagoes may have a weakened primacy effect.

We are now ready to test the city-country model for the largest city in the country, using the format of previous Figure 2. We expect the likelihood of primacy to decline as total population increases and a greater number of other relatively large and hence multifunctional cities is introduced. Figures 1 and 3 confirm this expectation: the ratios P_2/P_1 and also P_4/P_2 increase as total population becomes very large. Figure 4 shows P_1 graphed against P . The maximum conceivable population the largest city could have is $P_1 = P$, when the country consists of a single city. The theoretical curves in Figure 4 are based on Eq. 4 and 7. Countries of about 100 million are the only ones for which we would expect the data points to fall mainly in the zone delineated by the curves $n=0.9$ and $n=1.1$. For larger countries many points are expected to fall below that zone, as in Figure 2. For smaller countries, many points are expected to be above that zone, because of the primacy effect. The pattern of actual data points is fully as expected.

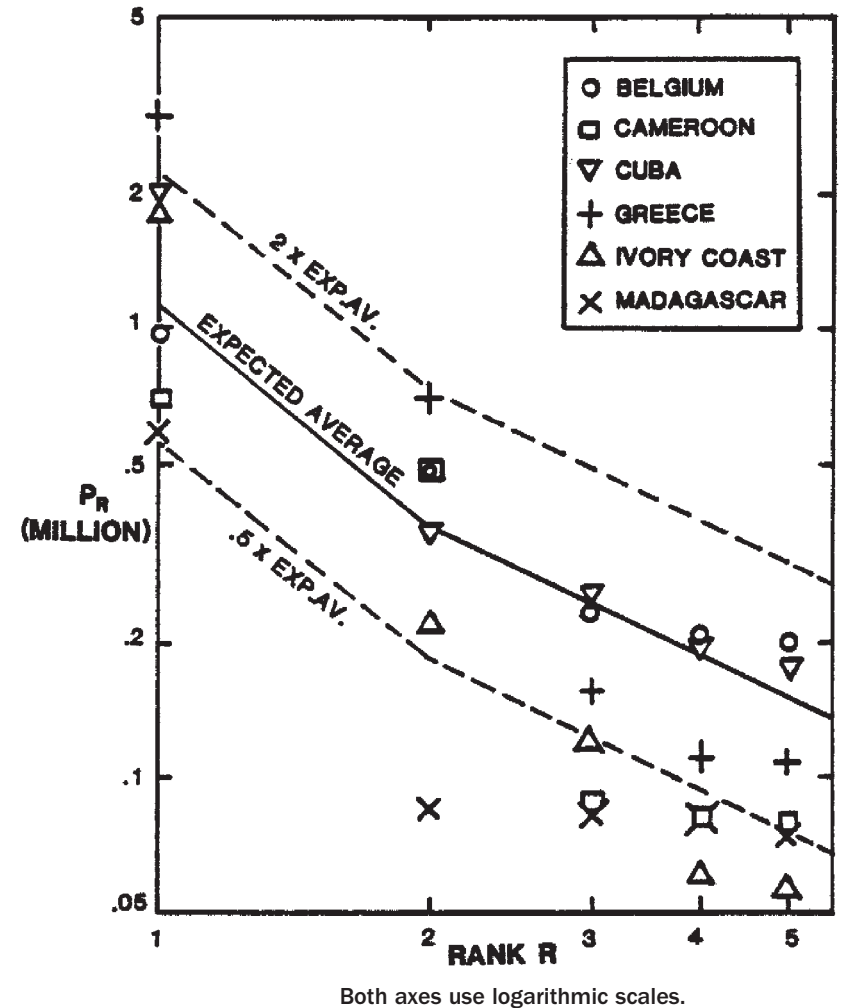
The previous Table 1 shows the median values of n for the largest cities, indicated as $n(P_1)$, for various ranges of country population. These values are close to 1.00 only for country populations over 10 million, reflecting an absence of primacy effect. As countries become smaller, $n(P_1)$ rises markedly above 1.0, indicating an increasingly strong primacy effect. For the world as a whole, $n(P_1)$ falls clearly below 1.0, and its value is, in fact, lower than those of $n(P_2)$ to $n(P_5)$. Close to 50 percent of the largest cities still are located in the zone between $n=0.9$ and $n=1.1$, but this figure is somewhat misleading, since the deviant points are preponderantly above this zone rather than spread equally on both sides.

PREDICTION AND EVALUATION OF CITY POPULATIONS

This is the payoff section. The city-country rule allows us to predict the world average populations of cities for a given country population. It then enables us to say whether cities in a given country are relatively large or small, compared to this world average. This is something the rank-size rule cannot do.

In terms of prediction, the best that rank-size rule can offer is to take the size of one city (maybe that of the second-largest, to avoid the primacy effect) and predict the rest. This is a weak peg on which to hang the whole scale, because the given city could be relatively small or large compared to all others. To guard against that, one would have to graph P_R against R and then look for a city that fits the average pattern—but that changes prediction into a *retroactive* search for a systematic relationship. In contrast, the population of the entire country is a much more stable comparison point, since it includes the populations of all settlements. Once the country population is given, the city populations are bound to fall in place to some extent.

Figure 5 – City populations in six countries with exactly 10 million population: actual and as predicted by city-country rule with primacy adjustment.



According to Table 1, the simplest form of the rank-size rule (that with $n=1.0$) applies on the average over a wide range of country populations (10,000 to at least 100 million). If we know the country population P , we could use $P = P_1 / nP_1$ (Eq. 4) to determine a theoretical value of P_1 , and then use $RP_R = P_1$ (Eq. 1) to calculate all other city populations. Two technical problems arise: 1) in the equational form above, P could be directly calculated from P_1 but not vice versa; and 2) the value of P_1 obtained does not include the primacy effect. These issues are addressed in Appendix B.

Figure 5 shows the resulting average expectation and actual data for six median-range countries with the same population: 10 million (plus or minus 0.20 million, according to *The World in Figures*, 1988). Both axes again use logarithmic scales. The dashed curves indicate a deviation from expectation by a factor of two. The median pattern of the six countries is somewhat steeper than predicted. The median P_1 (1.4 million) is very slightly on the high side. The median P_2 (0.4 million) is exactly as predicted. The median P_3 to P_5 are low by a factor of two.⁷

How do the individual countries fare? *Belgium* fits the expected average pattern very closely, and so does *Cuba*, except for an unusually large capital city. In *Greece* the two largest cities are unexpectedly large, while the lower-ranking cities are small.

If one used simply the rank-size rule, *Ivory Coast* would seem to satisfy it (apart from a strong primacy factor), since the points for second- to fifth-largest cities fall on a straight line (with a very steep slope corresponding to $n=1.9$). From the perspective of the city-country rule, however, all cities beyond the first are much too low. In *Madagascar*, the rank-size rule would lead to the impression that heavy primacy depresses the population of all other cities. The city-country rule adds the extra insight that even the largest city itself is rather small for a country of that population. In *Cameroon*, too, what appears a two-city primacy (akin to Greece) turns into two-city normalcy, plus unusually small third- to fifth-ranking cities. Thus we have to specify two types of primacy: relative primacy (compared to other cities) and absolute primacy (compared to the country as a whole).

CONCLUSION

The model presented pegs the entire rank-size pattern of cities to the total population of the given country. It thus leads to new ways to make worldwide comparisons, expanding the scope of the rank-size rule appreciably. The test with second-ranking cities in Figure 2 (and also with the third- to fifth-ranking cities) is successful in confirming the empirical validity of the simplest formulation of the rank-size rule ($n=1.0$), since the actual data points essentially scatter evenly around the $n=1.0$ curve at all population sizes.

⁷ This subset of countries considered covers the typical range of the ratio P_4/P_2 in Figure 1, but the median ratio of the subset (0.35) is below the general value of 0.5. The median ratio P_2/P_1 (0.21) is also on the low side compared to the average of countries of about that size in Figure 3, which is 0.33.

The city-country rule also leads to the development of some ways to deal quantitatively with the question of primacy. The test with first-ranking cities (Figure 4) agrees with the general model qualitatively. Although we still lack a quantitative model of primacy, at least we obtain an empirical expression for the worldwide average primacy factor as a decreasing function of the country population.⁸

On the basis of this information, it is now possible to develop an expected city population profile for individual countries of a given population, as illustrated in Figure 5. We can tell not only how cities stand compared to other cities in the same country (as the rank-size rule does) but also how any of these cities stand in comparison with similarly ranked cities in other countries of similar population. In particular, primacy can now be measured not only in comparison with other cities in the same country but also in a world context.

How satisfactory is the fit in Figure 5? Is it sufficient that Belgium, Cuba and (marginally) Greece agree with the expectations only within a factor of 2? Is the model failing in the case Cameroon, Ivory Coast and Madagascar where the deviation from expectation is even larger? It depends on the criteria for satisfactory. There is the absolute one: How close are the data to an ideal fit with the model? But there is also the relative one: Is the model doing better than competing models? In the latter respect, the answer is a definite "yes," because to our best knowledge no other models have been presented to tie city populations to the population of the entire country. In this sense, we have made considerable progress.⁹

⁸ Scholars used to dealing with empirical data fits may be concerned that the R-squares in Figures 2 and 4 are only around 0.8, while they are typically higher than 0.9 for rank-size fits. If so, what have we gained? The rank-size fits usually apply to a single country, while here we are dealing with all countries in the world. The surprise actually is that worldwide scatter is only slightly less than for individual countries.

⁹ The strong deviance from expectation in the case of the African countries in Figure 5 leads to a broader question. In the case of low urbanization all cities in a country of a given population are almost bound to be smaller. If so, can the model be expected to hold in earlier historical periods when urbanization worldwide was appreciably lower? Of course, organized countries in earlier times also tended to have smaller total populations (Taagepera 1997), but the question remains.

APPENDIX A: The Relation Between the Populations of a Country and its Cities

The Model for n=1

Consider first the simple rank-size rule corresponding to n=1:

$$R P_R = \text{constant} = P_1$$

where P_1 is the population of the largest city. Hence $P_R = P_1/R$. Assume that this rule applies literally to every settlement, down to the last isolated hut with a population of one person, with rank L : $P_L=1$. This is a natural cut-off point. Replacing P_R by 1 in the equation above yields $L=P_1$. This is a somewhat surprising result: If the simplest form of the rank-size rule applied to all settlements in the country, the number of these settlements would equal the population of the largest city. The total population (P) of the country is then the sum of all P_R , from $R=1$ to $R=L$. This summation can be approximated by the integral of $P_R = P_1/R$, from $R=1$ to $R=L$:

$$P = P_1 \int_1^L dR/R = P_1 (\ln L) = P_1 (\ln P_1),$$

which is our Eq. 4. Since $P_1=RP_R$, the population of the country can be connected to the population of a city of any rank, leading to Eq. 3: $P = R P_R \ln(R P_R)$. In particular, the formula for P_2 (as used in Figure 2) is $P = 2 P_2 \ln(2 P_2)$.

Readers who find the starting assumption of one-person settlements questionable need not worry. An unwarranted assumption would be signaled by the resulting dispersal of data or lack of agreement with the predicted curve. The naive model may not be as unrealistic as it might look. To be a fair approximation, the model does not really demand that the rank-size relationship apply literally down to individual isolated farms and other dwellings, because the latter account for a relatively small proportion of the total population. If we cut off at some P_L higher than 1, the total population of the country would be $P = P_1(\ln P_1 - \ln P_L)$. For $P_1=1$ million, using $P_L=10$ instead of $P_L=1$ would reduce the total population of the country only from 13.8 million to 11.5 million.

A better approximation to the summation would be obtained by integrating from $R=0.5$ on (rather than $R=1$). Compared to the actual summation, our equations underestimate the country's population by about $P_1/2$. Note that this correction works in the opposite direction to increasing P_L to higher than 1 and thus partly cancels it out.

The Model for Any Value of n

For the more general form of the rank-size rule (any n), where

$$R^n P_R = \text{constant} = P_1,$$

a population of $P_L=1$ would be reached for a rank $L=P_1^{1/n}$. Again, approximate the summation from P_1 to P_L by integration:

$$P = P_1 \int_1^{P_1^{1/n}} dR/R^n = (1-n)^{-1} (P_1^{1/n} - P_1) = P_1 (P_1^{-1+1/n} - 1) / (1-n),$$

which is our Eq. 7. In view of Eq. 2, the population of the country can be connected to the population of a city of any rank, leading to Eq. 6:

$$P = R P_R (P_R^{-1+1/n} - R^{n-1}) / (1-n).$$

For $R=2$ (as used in Figure 2), Eq. 8 results in $P = 2P_2 (P_2^{-1+1/n} - 2^{n-1}) / (1-n)$. For $n=1$, the value of P in the equations above is not defined. However, as n tends toward 1, the values of P tend toward values given by the earlier equations for $n=1$.

APPENDIX B: Estimation of P_1 from $P=P_1 \ln P_1$ plus primacy effect

One could determine P_1 from P graphically, using the $n=1$ curve in Figure 4. However, since that curve is almost a straight line (on log-log graph), $P=P_1 \ln P_1$ can be approximated by the equation of this line, which corresponds to

$$P_1(\text{theoretical}) = .256P^{925}.$$

The values of P_1 obtained from this approximation differ from those obtained from $P=P_1 \ln P_1$ only by at most 3 percent, for P ranging from 50,000 to 500 million. Even at $P=10,000$ or at 10 billion the discrepancy increases to only 7 percent. In conjunction with Eq. 1 this becomes

$$P_R = .256P^{925}/R.$$

This is the best average estimate of P_R on the basis of P , for R larger than 1. For the largest city, a primacy correction must be added, as follows.

Comparison of actual values of P_1 to those predicted by $n=1$ in Figure 4 leads to the following very approximate expression for the average primacy factor:

$$p = (\text{actual } P_1) / (\text{theoretical } P_1) = 0.5(10 - \log P),$$

where decimal logarithms must be used. Combining this equation with $P_1(\text{theoretical}) = .256P^{925}$ leads to an estimate of actual P_1 using decimal logarithms:

$$P_{1=p} P_1(\text{theoretical}) = .128P^{925}(10 - \log P).$$

For P larger than 10^6 , primacy factor is less than one, reflecting the "negative primacy effect" observed for very large countries. In such cases the population of the second-largest city may also have to be adjusted downwards so as to keep P_2 smaller than P_1 .

REFERENCES

- Alden, J. R. (1979). "A reconstruction of Toltec period political units in the Valley of Mexico." In *Transformations: Mathematical Approaches to Culture Change*, ed. C. Renfrew and K.L. Cook, pp. 169–200, New York: Academic Press, New York.
- Asami, Y. (1986). "Fitting the rank-size rule to legal cities." *Geographical Analysis*, 18:243–54.
- Auerbach, F. (1913). "Das Gesetz der Bevölkerungskonzentration." *Petermann's Geographische Mitteilungen*, 59–1:74–77.
- Beckmann, M. J., and McPherson, J. C. (1970). "City-size distribution in a central place hierarchy: an alternate approach." *Journal of Regional Science*, 10:1: 25–33.
- Beguín, H. (1979). "Urban hierarchy and the rank-size distribution." *Geographical Analysis*, 11:2:149–64.
- Berry, B.J.L. (1966). "Essays on commodity flows and the spatial structure of the Indian economy." RP III, Department of Geography, University of Chicago, Chicago, IL.
- Berry, B.J.L. and J. D. Kasarda. (1977). "Urban hierarchies and economic development." In *Contemporary Urban Ecology*, ed. B.J.L. Berry and J.D. Kasarda, pp. 386–400. New York: MacMillan.
- Chase-Dunn, C. (1985). "The system of world cities, AD 800–1975." In *Urbanization in the World Economy*, ed. M. Timberlake, pp. 269–272. Orlando, FL: Academic Press.
- Christaller, W. (1966). *Central Places in Southern Germany* (C.W. Baskin, trans.). Englewood Cliffs, NJ: Prentice-Hall.
- Dacey, M.F. (1966). "Population of places in a central place hierarchy." *Journal of Regional Science*, 6:2:111–24.
- Ettlinger, N., and Archer, J.C. (1987). "City-size distribution and the world urban system in the twentieth century." *Environment and Planning, A* 19:1161–74.
- Lösch, A. (1954). *The Economics of Location* (W.H. Woglau, W.F. Storper, trans.). New Haven, CT: Yale University Press.
- Mulligan, G. (1976). "Properties of a general hierarchical city-size model." *Geographical Analysis*, 8:4:395–405.
- Okabe, A. (1977). "Some reconsiderations of Simon's city-size distribution model." *Environment and Planning A*, 9:9:1043–53.
- Simon, H.A. (1955). "On a class of skew distribution functions." *Biometrika*, 42:425–40.
- Skinner, G.W. (1977). "Regional urbanization in nineteenth century China." In *The City in the Late Imperial China*, ed. G.W. Skinner, pp. 211–49. Stanford, CA: Stanford University Press.
- Stewart, J.Q. (1947). "Empirical mathematical rules concerning the distribution and equilibrium of population." *Geographical Review*, 37: 461–85.
- Taagepera, R. (1997). "Expansion and contraction patterns of large polities: Context for Russia." *International Studies Quarterly*, 41: 475–504.
- Taagepera, R. (2001). "Party size baselines imposed by institutional constraints: Theory for simple electoral systems." *Journal of Theoretical Politics*, 13:4:331–354.
- Vining, D.R. (1974). "On the sources of instability in the rank-size rule: Some simple tests of Gibrat's law." *Geographical Analysis*, 6:4:313–29.

- Vining, D.R. (1986). "Population redistribution towards core areas of less developed countries." *International Regional Science Review*, 10:1–45.
- The World in Figures*. (1988). Compiled by *The Economist*. Boston, MA: G.K. Hall and Co.
- Walters, P.B. (1985). "Systems of cities and urban primacy: Problems of definition and measurement." In *Urbanization in the World Economy*, ed. M. Timberlake, pp. 63–85. Orlando, FL: Academic Press.
- Zipf, G.K. (1949). *Human Behavior and the Principle of Least Effort*. Reading, MA: Addison Wesley.