A Note on the Rank-Size Rule and Future Urban Growth Patterns in Pakistan

GEORGE HEITMANN, WARREN C. ROBINSON
and
JEAN KAREN SALTER*

Urbanization is an integral part of the overall process of economic and social development. There is growing evidence that this is the case for Pakistan also. Pakistan in 1972 was nearly 27 percent urban and today is perhaps 30 percent plus urban. The probable distribution by size of place of this growth is of great importance for economic and social planning yet curiously it has received very little scholarly investigation as yet [7], is an exception and the present article should be viewed as an extension to Helbock's work.

In searching for some model or theoretical statement in the literature regarding what the future hierarchy of urban place in Pakistan will look like, Helbock is naturally lead to the "Rank-Size Rule" vs. "Primate City" question. After examining the 1961 and 1972 census data on urban places by size of place, he concludes that the "Rank-Size" distribution does not adequately describe the Pakistan urban hierarchy.

"In conclusion it may be said that the rank-size relationship clearly demonstrates that Pakistan's urban population is becoming more concentrated in a few of the nation's largest cities.... There appears to be a dearth of cities of the 100,000 to 250,000 population range and this fact also contributes to the lack of rank-size conformity" [7, p. 320].

Helbock is clearly correct regarding the 1961 and 1972 census data on urban places. Yet is this an adequate "test" of the rank-size rule as a predictor of the future urban hierarchy? Perhaps not for several important reasons. First, there appears to be some apparent confusion from which arises a conceptual misunderstanding of what the rule actually "predicts". The rank-size distribution constitutes an equilibrium or "steady-state" distribution which arises over time when the underlying forces making for urban growth are relatively constant in their effect and when the overall system is experiencing no

---

*The authors are associated with the Pennsylvania State University, USA.

external shocks. There is, thus, a direct analogy with the emergence of a stable age distribution from a constant regime of fertility and mortality rates. Berry has shown that, statistically speaking, the rank-size distribution will eventuate whatever the initial distribution from the working of Gibrat's "Law of Proportionate Growth" given enough time. Zipf and others stressed the evolution of an integrated social and economic system, a "nation", in which all places interacted with one another and formed part of a larger whole [1, 2, 11, 12].

Second, the system of cities which is the urban hierarchy in Pakistan is of very recent origin, dating in fact from Pakistan's founding in 1947. For prior to that time the cities in what is now Pakistan were part of the larger system of cities which made up the urban hierarchy in old British India. For this larger, mature hierarchy the rank-size rule fitted the urban hierarchy rather well prior to 1947 [6]. Following 1947, the Pakistan urban hierarchy begins in a totally disequilibrium state and there is every reason to think that the distribution will be some time in working out these disequilibrium forces. Stephen K. Lewis has shown that the most powerful single factor guiding the first 20 years of Pakistan's industrial pattern of growth was the fact of political separation from the larger Indian economic system and the consequent need for the Pakistan system to develop new capacities to meet its own needs [8]. In other words, from an initial post-Independence disequilibrium industrial pattern, market forces lead it towards a new industrial pattern reflecting its own national needs and resources availabilities. Why should this not also be true for the urban hierarchy? As a new equilibrium national economic pattern emerges and as this stabilizes, there is every reason to think an equilibrium demographic pattern will emerge as well. But it is important to see that this will require time to work out. How much time? We can not be sure but perhaps it will be useful to consider an illustrative example drawn from the experience of the U.S.A. during its own economic development.

To repeat, the best test of the rank-size rule would be through time and under conditions in which one could reasonably expect that "steady-state" growth was occurring. Let us look at data for the conterminous U.S.A., 1900-1970. The experience of the U.S.A. during this rather long time interval ought to offer at least some insights into the process under consideration and the availability of a good data series makes possible somewhat more formal tests of significance.

THE DATA USED

The Concept of Standard Metropolitan Statistical Area was first employed in the 1950 Census of the U.S.A. and, with modifications, has been used in each census since then. Conceptually it stems from the metropolitan district notion, that the real socio-economic unit which should be measured as the city does not end at the political boundaries but instead encompasses suburbs and satellite cities and even open country nearby areas. Thus, a metropolitan "place" is not a single urban place but a cluster of many places plus other population as well [4, 9]. This must be considered as representing the highest state of urbanization and the urban hierarchy's first subdivision should be mertropolitan places.

We then reconstructed an urban hierarchy in the U.S.A., 1900 to 1970 in which the top four size groups are SMSA's, while the others are urban places located outside the dominance zone of any of the SMSA's.

A word or two is in order about this series. The SMSA data do not go back before 1950 and the present 1900-1940 SMSA's are "retrojections" of the 1960 Census-defined SMSA's [5]. These retrojected SMSA's were themselves necessary before the non-SMSA urban places could be derived. Some limitations are also inherent in the method. SMSA's are still a relatively crude statistical unit made up of counties. Changes in the boundaries of the counties and also the SMSA's have been ignored. Also, the SMSA's include suburbs while the non-SMSA urban places do not, thus making for an under-statement of the urban population in these categories [3, 10].

Attempting to rank-size order all 3,000 plus urban places in the U.S.A. for each census year, 1900 to 1970, while possible would also be laborious. Working with grouped data is clearly quicker and makes the results easier to interpret. Thus, we used the nine size of place categories shown in Table 1 and assume that the distribution of the total urban population contained within the hierarchy also follows the rank-size rule. That is, we may substitute the rank of the class for the R in the rank-size equation, and for P_r the population of the whole category, and still obtain the constant.

A simple test of the rule's applicability to these data is possible. If the pure $R (P_r) = M$ holds then whatever the size of the largest place (or the population contained in the largest category), the second largest place (or group) should have half this population, and so on for the third, fourth, and so on places (or groups). From this, in turn, one can derive the percentage distribution of the population contained in the entire hierarchy size of place groups.

Putting it another way, we have nine categories and if the rank-size rule is correct, then the probability of a randomly selected city being in category one is 0.3546 and so on for the other categories. These probabilities or percent distribution can be viewed as "expected" if the rank-size rule is correct.

The derivation of these "expected" values is shown below:

Rank	Population in Group	Percentage of Total Population
1	1.00	35.46
2	.50	17.73
3	.33	11.70
4	.25	8.87
5	.20	7.09
6	.18	5.67
7	.14	4.96
8	.12	4.61
9	.11	3.90
		100.00

Our use of categories is unorthodox but not unique. Some of the richness of the entire distribution has obviously been lost by grouping but what has been gained is computational ease and also much easier results to interpret and analyze. The categories could obviously be juggled in such a way as to force a fit in any one year but this was not done. However, it is also obvious that such an accommodation would only work for one year. Since the categories remain unchanged over the entire period 1900-1970 no amount of juggling of the boundaries between categories would guarantee a fit for all years.

SIGNIFICANCE OF THE RESULTS

Table 1 presents this "expected" percent distribution of total urban-metropolitan population compared to the actual distributions, 1900 to 1970. One's casual impression from looking at the table is that some of the actual fit the "expected" better in some years than in others. Fortunately our approach lends itself to more rigorous tests of significance.

First, Table 2 presents the correlation coefficient (r) relating "expected" values (X) to the observed values (Y). Thus, in $Y = a + bX$ we are, in effect, attempting to predict the actual values on the basis of the "expected" values and r is a measure of the success of this prediction effort. As may be seen in Table 2 these r values are all extremely high and, moreover, show a definite upward movement over the period 1900 to 1970. Our expectation is also that a should be zero or a value not significantly different from zero and this proves the case from 1930 to 1970. In 1900 to 1920 inclusive a was, in fact, significantly different from zero. Suggesting from this point of view also that the rank-size hypothesis became a better and better proposition as we move from 1900 to 1970. Table 2 also reports the b values. If $a = 0$, and $r = 1$, then $b = 1$ also. The b values are also high and the standard deviation of their own distribution indicates a strong clustering around the mean value.

Thus, correlation-regression anlaysis indicates that the "expected" and actual distributions are statistically highly similar.

A second statistical test is also possible. The statistic:

$$\Sigma \frac{(\text{observed-expected})^2}{\text{expected}}$$

follows the $\times^2$ distribution. In this case, with 8 degrees of freedom (or n-1), the critical value $\times^2$ are 13.362 (at the 10 percent level of significance) and 15.507 (at the 5 percent level of significance). Using this test there does seem a tendency for the computed $\times^2$ statistic to become smaller over time 1910 to 1970. The 1960 and 1970 values are well below the critical levels while 1910 to 1950 are not. This is reasonably consistent with the correlation-regression results, indicating an increasing "goodness of fit" through time.

CONCLUSIONS

Using grouped data for a modified urban-metropolitan hierarchy which treats Standard Metropolitan areas as single urban places, we find that in the U.S.A. from 1900 to 1970 there was a growing movement of the observed distribution towards the distribution to be expected (or predicted) by a simple

Table 1

Expected and Observed Distribution of U.S. Urban-Metropolitan Population by Size of Place Categories

Rank	Expected	Observed							
		1900	1910	1920	1930	1940	1950	1960	1970
1	35.46	16.76	25.33	23.62	31.20	30.51	30.85	30.31	31.64
2	17.73	13.33	5.17	8.44	7.91	7.62	9.10	13.23	17.23
3	11.70	10.35	13.32	14.11	10.37	13.24	15.37	15.56	13.29
4	8.87	19.40	20.52	19.58	19.38	17.87	16.63	15.32	15.07
5	7.09	11.50	8.57	9.67	9.03	8.72	7.64	8.12	5.98
6	5.67	6.65	7.56	6.83	6.49	6.17	5.69	5.06	5.07
7	4.96	8.43	7.48	7.01	6.58	6.61	6.38	5.33	4.91
8	4.61	6.67	6.00	5.48	4.83	5.06	4.64	3.88	3.79
9	3.90	6.91	6.04	5.26	4.20	4.20	3.69	2.99	3.02

rank-size rule. By 1970 it was statistically a very good fit, using both correlation-regression and χ^2 test of significance.

Thus, the rank-size interpretation of our results would be that the urban-metropolitan hierarchy in the U.S. was not a "mature" well-integrated socio-economic system in 1900 but increasingly became such with the passage of time. In other words, the development of an increasingly mature system also lead to the emergence of the expected urban distribution.

Finally, we suggest that in a scant 25 years since Independence a mature fully-developed urban hierarchy simply has not had time to evolve in Pakistan. But it took much longer than that in the U.S.A. under more stable conditions. In any case, the planners may very well be lead astray if they look only at the "Primate" cities of Karachi and Lahore and fail to allow for developments in the other rungs of the rank ordering of urban places.

Table 2

Measures of Statistical Significance of the Relationship of the "Expected" and Actual Distributions of Urban-Metropolitan Places in the U.S.A., 1900-1970

	b	SD_b	r	χ^2
1970	0.891	0.092	0.964	9.259
1960	0.806	0.115	0.936	11.875
1950	0.763	0.148	0.890	14.204
1940	0.714	0.160	0.860	14.571
1930	0.733	0.175	0.846	15.611
1920	0.491	0.161	0.755	15.641
1910	0.499	0.192	0.701	15.036
1900	0.282	0.135	0.619	11.059

REFERENCES

1. Berry, J.L. Brian and Allen Pred. *Central Place Studies.* Bibliography Series No. One, Philadelphia. Regional Science Research Institute, 1965.

2. —————. "City Size and Economic Development". in: Leo Jakobson and Ned Prabcash. *Organization and National Development.* Beverly Hills: Sage Publications, 1971.

3. —————. Peter G. Goheen and Harold Goldstein. *Metropolitan Area Definition: a Re-Evaluation of Concept and Statistical Practice.* Bureau of the Census, U.S. Department of Commerce, Washington, D.C., Working Paper No. 28, 1968.

4. Bogue, Donald J. *The Structure of the Metropolitan Community.* Michigan Press, 1949.

5. —————. *Population Growth in Standard Metropolitan Areas,* 1900-1950. Housing and Home Finance Agency, U.S. Government Printing Office, 1953.

6. Davis, Kingsley. *The Population of India and Pakistan.* Princeton Univ. Press, 1950.

7. Helbock, Richard W. "Urban Population Growth in Pakistan: 1961-1972". *The Pakistan Development Review.* Vol. XIV. No. 3, Autumn 1975.

8. Lewis, Stephen K. *Industrial Development in Pakistan,* 1947-1965. Allen and Unwin, 1970.

9. McKenzie, R. *The Metropolitan Community.* McGraw-Hill, 1933.

10. Robinson, Warren C. *Metropolitan and Urban Growth in the U.S.A.,* 1900-1960. Institute of Research on Land and Water Resources, The Pennsylvania State University. Report No. 58, 1968.

11. Steward, John Q. "Empirical Mathematical Rules Concerning the Distribution and Equilibrium of Population". *Geographical Review.* 37, 1, 1951.

12. Zipf, George K. *Human Behavior and the Principle of Least Effort.* Addison-Wesley, 1949.