

Reprinted from ECONOMIC DEVELOPMENT AND CULTURAL CHANGE
Vol. 33, No. 1, October 1984
© 1984 by The University of Chicago. All rights reserved.
Printed in U.S.A.

Statistical Determinants of the Population of a Nation's Largest City*

Lee De Cola
University of Ibadan

I. Introduction

In spite of the fact that urbanization on a world scale is the dominant spatial demographic phenomenon of the twentieth century, social scientists manifest a striking lack of unanimity about the nature, causes, and possible future of the urban revolution. What is clear to all observers, however, is that the continuing growth of cities is an extremely uneven phenomenon: while an overall urbanization rate for a nation may reflect the general growth of all urban places, classes of cities may experience a wide range of rates, with individual places showing even more heterogeneous histories.¹

This paper is an attempt merely to understand the tip of the huge and growing urban iceberg. The question I wish to answer is, What are the important statistical determinants of the population of a nation's largest city? The very asking of this question calls for a statement of my normative and scientific reasons for the investigation. On the one hand I shall make no judgment whether a given city may, with respect to some set of criteria, be too big or small. Nevertheless, it will be clear below that some number-one cities are relatively very much larger than others even when their characteristics and national environment are taken into account. On the other hand, there are sound scientific reasons (such as intriguing patterns in the data) for focusing analysis on the upper tail of the national urban size distribution.² Furthermore, because a country's largest city often contains one-third or more of its urban population, there are strong policy motivations for this investigation.

The analysis presented below has four goals. First, the explication of links between certain variables and city population is intended to be a contribution to a theory about what factors influence urban size in general and the size of CITY1 (the largest city in the nation) in particu-

lar.³ The approach adopted here is quite straightforward: to introduce a set of predictor variables into a regression model to see how well they predict LOGPOP1 (the logarithm of the population of CITY1). If urban science is at all well developed, there should be few surprises when this is done.

A second goal of the analysis is to examine the residuals from the model in detail to attempt to discover why a given city is larger or smaller than the model predicts. A medical analogy may be suggestive here. When a patient visits a doctor the first questions the doctor asks relate to symptoms. Often the symptoms are not important in themselves but only as indicators of some underlying process not readily apparent to the patient. Hence the "primacy syndrome" receives more attention than it may deserve in itself because it may point to underlying problems in the national distribution of population.⁴ The analysis reported below will add another diagnostic tool to the study of primacy by generating regression residuals that may be treated as measures of relative urban concentration.

Related to the second goal is a third, to examine the outliers among the residuals for anecdotal insights into the processes generating urban growth. Just as the pattern of residuals throws light on the process of urban concentration, so may the extreme cases show where the processes break down because of "pathological" factors.⁵

A final goal is to direct this analytical light on Africa, a continent that is still relatively unurbanized but is experiencing the world's highest rates of urban growth, and the continent where colonialism continues to play a visible role in urban processes.⁶ Until recently Latin America and Asia have been the favorite laboratories for students of rapid urbanization and large cities, but improved data—and rising concern—demand that the regions of Africa, with their richly varied histories and amazingly diverse current developmental paths, receive much more attention than they have in the past.

Because the word will figure in the analysis that follows, I must begin with a coherent statement of what precisely shall be denoted by "primacy." There seem to be at least three meanings. First, to follow the Latin etymology, CITY1—whatever its size—experiences primacy by being the primate settlement in the national set. Second, there are occasionally cities other than CITY1 (like Jerusalem, Rio de Janeiro, and Peking) which occupy a preeminent place in the urban functional hierarchy and therefore partake of primacy as well. Third, the word has been used in the urban geographical literature more selectively to connote the "excessive" size of CITY1 often observed in the national urban size distribution. The question naturally arises, Excessive with respect to what? Usually the criterion is the linear prediction of a rank-size line or the mean value of some ratio. I intend to expand the set of urban concentration measures with a new and arguably more robust

measure of the relative size of CITY1. I shall attempt, however, to avoid the use of the more value-laden term "primacy."

The organization of the paper is as follows. Section II develops a theoretical treatment of the extreme value problem as applied to urban size and presents an intuitively satisfying path-type model to predict urban population for any city. Section III describes the data used to calibrate the model and gives a brief picture of the international database. Section IV presents the principal findings: regression coefficients, residuals, outliers, and regional analysis; while Section V sets forth a few conclusions, policy implications, and suggestions for further research.

II. Theory and Model

Models predicting urban size have constituted a lively field of research during the twentieth century. Although the literature is large and varied,⁷ much of the work may be grouped under four major headings. First and perhaps oldest are systemic studies that attempt to predict the population of a city from its rank among the cities of a set, the theory being that the set forms a system, each of whose members has a somewhat determinate place in the ranking. Second are hierarchical models, such as those of economic central place theory and political-administrative organization, that predict size from function and service areas. Third and most diverse are stochastic models, among which are random splitting, Markov processes, and the lognormal process of growth. Finally there are more empirical studies of primacy and regional analyses, particularly in Latin America and Asia.

A full discussion of this broad field of research would take me much too far from the task at hand. It seems clear, however, that the most parsimonious model, the one requiring fewest assumptions, is the lognormal, which argues that $POP(i)$, where i is a given city (i need not be the rank), is a random variable whose logarithm is the normally distributed outcome of a stochastic birth and growth process. With this in mind, let us examine the global data.

Let us begin by imagining that we have compiled a list of all the settlements in the world, the problem of what determines a settlement having been solved.⁸ One way to understand the statistical distribution of urban sizes is to rank order this list of settlements by population. A recent United Nations study compiled data on all the cities of the world with an estimated population greater than 100,000 and arrived at the size distribution shown in table 1. If we assume that there are 8,000 cities in the world, then table 1 may be regarded as giving relative cumulative frequencies for the top 21%. A simple linear regression model allows us to estimate the mean and standard deviation of a lognormal distribution that predicts these percentages ($R = .9997$).⁹ The implication of this finding is that at least the largest cities of the

TABLE 1

SIZE DISTRIBUTION OF THE WORLD'S CITIES WITH POPULATION OVER 100,000 IN 1975

Population Interval	Number of Cities	Cumulative Number	Cumulative Percentage Assuming 8,000 Cities
Over 4,000,000	30	30	.4
2,000,000-3,999,999	48	78	1.0
1,000,000-1,999,999	107	185	2.3
500,000-999,999	227	412	5.1
250,000-499,999	441	853	10.7
100,000-249,999	802	1,655	20.7
Total	1655		

SOURCE—United Nations, *Patterns of Urban and Rural Population Growth* (New York: United Nations, 1980), p. 48.

world may be regarded as the tail of a lognormal distribution with a mean of 9.928 and a standard deviation of 1.965 (using natural logarithms).

The consequences of this assumed distribution are rather interesting. The model predicts that the largest city in the world should have a population of about 30,000,000 and the smallest (a very unreliable prediction for this model) a population of about 15, while the median settlement should contain about 20,000 people. But because the model predicts only 500 towns with a population less than 1,000 (6.2% of the settlements), it seems that the distribution applies to that part of the national population which may be called "urban" and not to all possible settlements. It is encouraging therefore to note that the total population residing in the 8,000 cities of the model is about 1.1 billion, not too far from current estimates of the world's urban population.¹⁰

From an empirical standpoint this discussion suggests that, whatever the underlying determinants of urban population, the lognormal model works quite well on a world scale and that the logarithmic transformation of population is justified (in addition to the obvious fact that it reduces the unmanageable skewness of the variable). Another feature of the transformation, which I will make much use of below, is that regression residuals of a logarithmically transformed dependent variable are the logarithms of the ratio of actual to predicted values.

Not much work has been done to model the world city population distribution because there is as yet no consensus on what kind of world system of cities exists—or whether one exists.¹¹ The discussion above is offered merely as a suggestion that, on a world scale, the lognormal model fits the data. But this is no proof of the existence of systemic processes. Except for the very largest cities, which certainly are becoming enmeshed in an ever-tighter network of financial, transportation, and communication links, the world distribution remains more

described than analyzed. Much attention, however, has been directed at urban populations for various partitions of the earth's surface—usually nations, but sometimes larger or smaller units such as continents or states.¹² Here again, the lognormal distribution usually fits the data, but the possibility of systemic relationships among the cities of a geographically connected subset has led to a great deal of work on “rank-size” models in which the size of a city is related to the number of cities larger than it.¹³

The focus of the present research is on the tip of the tail of the urban size distribution within a nation. The simplest model to explain POP1 would be based on the assumption that settlements are distributed randomly over the earth's surface and that nations are random partitionings of this surface.¹⁴ This model suggests that the extreme values of a nation's urban size distribution are simply the maximum values from spatially connected random samples. For example, if we assume 8,000 settlements with a given size distribution, then the expected maximum value for any country should be a function solely of that country's area. So the USSR should (probabilistically) contain the world's largest city because it represents the largest sample, followed by Canada, China, and so on.¹⁵ But although I have found that the zero-order correlation between area and LOGPOP1 (the population of CITY1) is .39, the very largest cities are found in such “unusual” places as Japan, Argentina, and France as well as the more extensive countries, suggesting the obvious fact that there are other important factors determining extreme values.

This problem—that the sizes of the largest cities may be in part the realization of a random process but that their spatial distribution is not random—suggests that we consider what specific factors determine the population of *any* city within a nation, accepting the fact that international differences reflect the influences of national characteristics. Before returning to a discussion of maxima, therefore, let us consider what these factors might be. This discussion is based on the theory that a model that predicts city size for any city should also work for CITY1, although it would have to be modified to take into account prior factors resulting in CITY1's being the largest.

I consider spatial variables first. National land area has already been mentioned as a pure spatial variable, but other factors, such as location within the national space, with respect to natural features, and in relation to other cities, should be included as well. Furthermore, because cities seem to be born, grow, and even die as part of what may be viewed as a self-organizing system—for if it is a system it certainly is not wholly exogenously organized—then such factors as the shape of a country might be significant determinants of location and, by implication, of size.¹⁶

Key socioeconomic variables (including political and technologi-

cal factors) constitute a second set of determinants of urban size. At the national level it would be expected that the rate of national population change is an important determinant of the rate of change, and therefore of the magnitudes, of the population of national cities. Economic development, as well, is known to have important but apparently nonlinear effects on urbanization and concentration, so that it is necessary to include at least one index of the extent to which the national environment is economically developed and technologically integrated.¹⁷ Included in this second group should also be those characteristics of the city itself that represent its economic and political role in the national urban system. At the minimum we should know whether the city is a national or regional capital and its place in whatever transportation and communication network may exist. Naturally, many other national and city indices could be added to a comprehensive model, but these factors will do for a start.

The third set of variables are the demographic factors likely to be of importance in predicting the shape of the national urban size distribution and especially the conformation of its upper tail. To borrow from the simple world model developed above, we seek factors that will be associated with three key parameters of the distribution. If, as seems likely, the national urban size distribution has a lower threshold, then the extent to which the nation is urbanized could stand for this lower limit.¹⁸ Next, the location of the distribution (its mean) will reflect the total national population, on the theory that big countries have big cities. Finally, the skewness of the distribution (which for the log-normal model is sensitive only to the standard deviation of the corresponding normal distribution) may be represented by the tendency of the urban population to be concentrated in the biggest cities.¹⁹ The factors that give a larger positive skew to the distribution not only are related to urbanization but also are reflected particularly in the sizes of the other largest cities. This suggests our examining the extent to which the population of the class of a nation's largest cities may be a determinant of the population of any one of them. If, for example, CITY2 is big then CITY1 will of course be bigger, but how much bigger is a reflection of the extent to which the national population is concentrated in the very largest city. (Another demographic index widely used in predicting the population of a city is its rank in the national distribution. As my attention will focus only on the population of CITY1, this index is only used implicitly here.)

Let us now take the factors predicting city size and organize them into a recursive system with causal implications. The eight concepts derived from the discussion above have 15 potential links. One way of interrelating these concepts is shown as a path-type model in figure 1. Each variable is causally related, directly or indirectly, to every other variable through a network of arrows, with the sign on each arrow

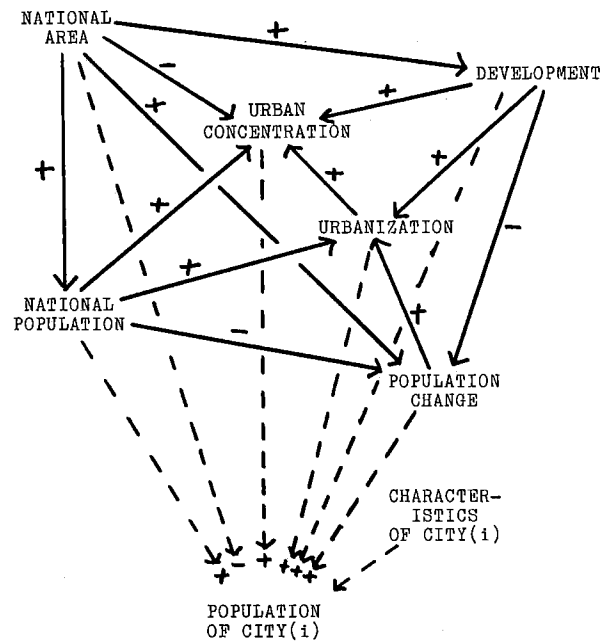


FIG. 1.—Hypothetical causal linkages among concepts

representing expected sign of the partial correlations given all prior variables. Although space does not permit the discussion of the full recursive system, at least three of the arrows deserve comment. The negative link between population and population change reflects simple negative feedback in which the mass of humans depresses further growth given fixed area. The negative link between population change and development reflects a crude assumption that development seems to be a shift from the production of people to the production of things. Finally, area negatively influences concentration by permitting large competing centers.

Three absent arrows also require explanation. No link is shown between national population and development because the relationship is likely to be nonlinear given constant area: in the early stages of population growth additions to national population will contribute to development, but it seems likely that really dense population concentrations, even on the national scale, will depress economic growth. Second, area is not tied to urbanization mainly because the demographic impact of space is expressed through the link to concentration, which is expected to be more sensitive to the availability of other megalopolitan sites than would be the development of cities per se. Third, there is no link from population change to concentration mainly because the present evidence suggests that population growth is quite

unevenly distributed among classes of cities.²⁰ In any case, the indirect link via urbanization seems adequate.

The model is quite general. For example, if $POP(k)$, the population of city of rank k , were at the focus the model would represent a rank-size analysis, correcting for city characteristics. In the present case, however, the diagram links key factors that may be operationally defined in a variety of ways to explain $POP1$.

The coherence of the model—the way it includes and interrelates the key variables—suggests that it should be successful in developing a straightforward way of predicting $POP1$, and at the same time reveal the signs and magnitudes of the linkages between the other indices. This model will inform the analysis below. It illustrates why I argue that the size of a nation's largest city is symptomatically useful in revealing the underlying urbanization and developmental processes.

III. The Data

The multivariate model discussed above was tested using data assembled from several recent sources, issued mainly by the United Nations. Although there is a great deal more detail about each country and the cities in the full database I have been using (such as populations and locations of each country's four largest cities, dates of settlement, and a number of explicitly political variables) the present model focuses on the indices most likely to have some explanatory power in predicting $LOGPOP1$. Descriptions, mean values, and sources for the variables are presented in table 2, while intercorrelations are shown in table 3.

All of the variables are operationalizations of the concepts from the model. $LOGLENGTH$ is, along with $AREA$, a crude measure of the shape of the national space. $LOGENERGY$ is a measure of economic development whose correlation with another measure, the logarithm of per capita GNP, is .92. $PCTURBAN$ is an estimate, according to each country's definition, of the percentage of the national population residing in urban areas. $LOGPOP2$ is used as a measure of concentration, for it represents the maximum value of the rest of the urban size distribution.

The sample of 126 nations includes every country on which the United Nations compiles data and which has at least one city with a population greater than 100,000 (see App. for a list of these cities). Because the spatial distribution of the largest cities is so far from random, the full database includes a complete rank-size distribution only for the first 19 cities in the world. Nevertheless, the countries represent about 97% both of the world's total population and of its inhabited area.

Because of this comprehensiveness, the meaning of the word "sample" is debatable here. In such cross-national research one must

TABLE 2
VARIABLES USED IN THE REGRESSION MODEL

Variable	Mean	Definition	Source
LOGPOPI	2.95	Log(base 10) of population of largest city in country in thousands, 1975 or latest	1
LOGAREA	2.45	Log of area of country in thousands of square kilometers	2
LOGLENGTH	1.01	Log of radius of circumscribing circle in hundreds of kilometers	3
LOGPOP	3.94	Log of national populations in thousands, 1976 or latest	2
PCTCHANGE	2.17	Percent annual change in national population, 1960-70	2
LOGENERGY	2.73	Log of energy consumption per capita in kg coal equivalent, 1975	4
PCTURBAN	40.1	Percent of population in urban places (own country's estimate)	2
LOGPOP2	2.43	Log of population of second largest city in thousands	1
PORT	.548	Dummy: 1 if seaport	3
CAPITAL	.833	Dummy: 1 if capital of country	5
METRO	.579	Dummy: 1 if city data are for metropolitan area	1

SOURCES.—(1) Mainly *U.N. Demographic Yearbook, 1977* (New York: United Nations, 1978); at least two other sources were consulted to confirm relative populations of cities.

(2) UN Conference on Trade and Development, *Handbook of International Trade and Development Statistics* (New York: United Nations, 1979).

(3) *The Times Atlas of the World* (Boston: Houghton Mifflin Co., 1967).

(4) *United Nations Statistical Yearbook, 1976* (New York: United Nations, 1977).

(5) Arthur S. Banks, *Political Handbook of the World* (New York: McGraw-Hill Book Co., 1977).

face the question whether a collection of virtually all the countries of the world constitutes a sample in the statistical sense. On the one hand, the collection is the full population of nations (minus a few small countries and city-states) and therefore questions of statistical significance are not strictly meaningful. I choose, on the other hand, to regard each nation as the outcome of a long experiment that could have turned out quite differently given the contingencies of history. Consequently the usual tests of significance remain valid and useful in that their associated statistics may be regarded as weights that reflect the importance of each variable to the model. As a result of this argument, I shall refer to the 126 cases as a sample both because they are technically not all of the countries of the world, or all possible realizations of the national experiment, and because just as a set may be regarded as a subset of itself, so may a population be regarded as a sample.

Nevertheless, the question of spatial autocorrelation must be acknowledged. This problem, like its temporal counterpart, can lead to underestimation of the variance of the regression coefficients. This is a

TABLE 3
ZERO-ORDER CORRELATIONS AMONG VARIABLES IN THE MODEL

	LOGPOI	LOGAREA	LOGLENGTH	LOGPOP	PCTCHANGE	LOGENERGY	PCTURBAN	LOGPOP2	PORT	CAPITAL
LOGAREA	.39									
LOGLENGTH	.43	.96								
LOGPOP	.81	.54	.54							
PCTCHANGE	-.27	.12	.10	-.20						
LOGENERGY	.45	-.06	-.04	.10	-.51					
PCTURBAN	.41	-.15	-.11	-.01	-.43	.81				
LOGPOP2	.85	.39	.41	.81	-.25	.44	.36			
PORT	.11	-.05	-.01	-.05	-.01	.10	.19	-.00		
CAPITAL	-.25	-.19	-.20	-.30	.09	-.16	-.14	-.42	-.19	
METRO	.13	-.10	-.07	.01	-.04	.05	.16	.08	.19	-.21

further reason for treating such estimates as the t -value as weights rather than as formal test statistics. But because national borders do enforce a fair degree of closure, and because the residuals from the model show such a great deal of regional diversity, I assume that a test of spatial autocorrelation would be negative.²¹

A less complex problem with the data is reliability. The dependent variable, POP1, is extremely difficult to measure. It is, like all urban populations, subject to underestimation resulting from difficulties of boundary definition and census undercounting. POP1 is also an object of some symbolic significance to national citizens and therefore it is likely as well to be overestimated. The fact that these two biases tend to be in opposite directions does not really lessen our uncertainty. I have attempted to reduce this uncertainty both by including a dummy variable representing whether the urban population figures given are for the city or for the metropolitan area, and by checking the population figures against three or more sources. But as these sources must to some extent draw on one another, the cross-checking verifies magnitude more than absolute size.

The independent variables, as well, suffer from the unreliability that plagues international data. With respect to the models built on such data, one expects the usual two kinds of error, type 1 and type 2. The second results from "noise," and although I expected it to predominate, the clarity of the results below suggests that there is surprising consistency in the data in spite of their diverse sources and conceptual foundations. Error of type 1, on the other hand, results from bias, and it is quite likely that to some extent the producers—and especially the compilers—of the data render them "consistent" by the implicit use of models such as that in figure 1. If this results in significant bias, then it is not surprising that we sometimes find what we expect to find. Until reality thwarts us, the best answer seems to be to include as wide a range of indices in our models as possible. This is further justification for my use of spatial and political factors in the model: either objective or discrete, they are in any case explicit.

IV. The Analysis

The model tested here predicts the logarithm of the population of each country's largest city using the ordinary least squares regression equation

$$\text{LOGPOP1} = \mathbf{Xb} + u,$$

where \mathbf{b} is a vector of coefficients estimating the marginal effects of a row vector \mathbf{X} of 10 independent variables, and u is an error term presumed to include the effects of excluded variables plus random disturbances.

TABLE 4

REGRESSION COEFFICIENTS

Variable	Coefficient	<i>t</i> -Statistic	Contribution to R^2 *
INTERCEPT	-.762	-3.43	...
LOGAREA	-.192	-2.21	.0051
LOGLENGTH	.460	2.41	.0061
LOGPOP	.632	9.13	.0875
PCTCHANGE	.0530	2.39	.0060
LOGENERGY	.0862	1.91	.0038
PCTURBAN	.00662	4.58	.0220
LOGPOP2	.155	2.24	.0053
PORT	.0869	2.29	.0055
CAPITAL	.171	3.11	.0102
METRO	.0690	1.81	.0035

NOTE.— $R^2 = .879$; adjusted $R^2 = .869$.

*The amount by which R^2 would be reduced if that variable were removed from the regression equation.

The estimates of the coefficients and their *t*-values are presented in table 4. The value of the multiple correlation coefficient suggests that the model works quite well in that it has reduced the variance of the dependent variable by 88%. Each of the important coefficients would, for a true random sample, be regarded as statistically significant, their *t*-values being a reflection of the marginal contribution of each variable to the explanatory power of the model. Each estimated coefficient will be discussed in turn.

LOGAREA is a negative determinant of the dependent variable because of the process discussed above whereby a large national space permits the evolution of competing urban centers. The (relatively) largest cities dominate the smallest countries.²² It is also interesting to note that, in spite of the collinearity between this variable and LOGLENGTH, the marginal influence of each is sufficiently strong to allow reasonably good estimates of their separate (and opposite) effects on the dependent variable.

The LOGLENGTH coefficient is, as previously remarked, probably a reflection of the process by which large cities often grow in eccentric regions where the competition from other centers is weaker.²³ As part of the overall research program I generated computer-drawn maps of each country in the sample. The pictures clearly show the eccentricities of the major cities. Furthermore, my ancillary analysis of the distance separating nearest neighbors among the four largest cities suggests that their spacing is wider than a random process would predict.²⁴

LOGPOP is clearly the most important predictor of LOGPOP1, and this variable is largely responsible for the high *R*-squared of the prediction. Furthermore, the magnitude of this effect is a clear valida-

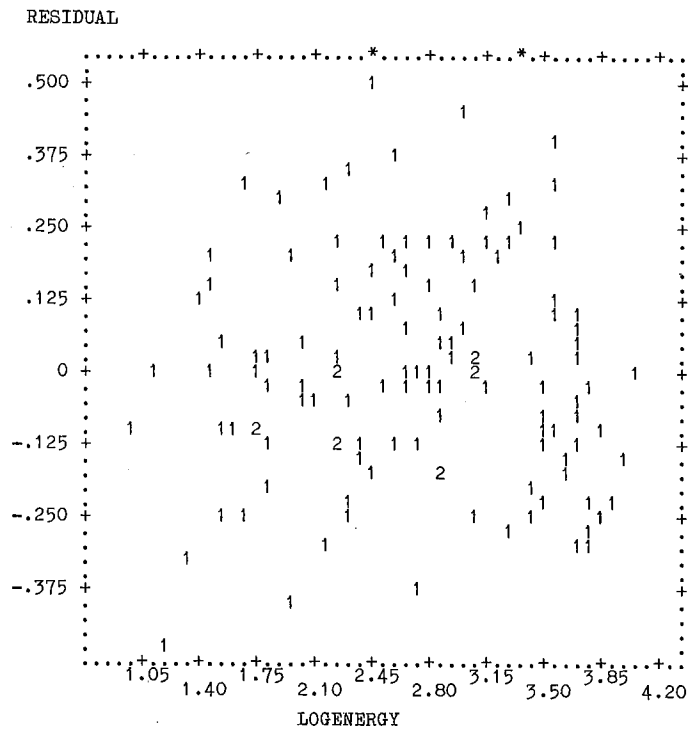


FIG. 2.—Scatter diagram: residual \times LOGENERGY

tion of the theory that a nation's cities do form a system in that their largest values strongly reflect the population of the country in which they are found.²⁵

It is clear from the coefficient for PCTCHANGE that the size of CITY1 to some extent reflects the growth rate of the nation. In a similar analysis of urban growth rates the United Nations study cited above found that the rate of change of the national population was by far the most important determinant of city growth. Just how important this link is may be seen from the model's implication that a 1% difference in a country's growth rate results in a 13% larger value for POP1. The model suggests that natural increase plays a heavy role in generating urbanization and urban concentration.

The finding for LOGENERGY is perhaps the most interesting. First, as expected, the size of CITY1 does to some extent reflect the level of economic development of a country. But, second, the effect is not very strong and may even be nonlinear. In this regard the scatter diagram of the residuals from the model versus LOGENERGY (fig. 2) shows that all 20 of the largest residuals are found among those countries where energy consumption per capita is between 50 and 4,000

kilograms of coal equivalent, while none of the 31 countries outside this range had such large residuals. This finding is elaborated on below.

The index of urbanization, PCTURBAN, is the second most important determinant of LOGPOP1. This finding is not surprising, but even for this variable nonlinearity in the residuals suggests (if I may leap to a dynamic argument) that once a nation is about 50% urbanized the relative size of its largest city may cease to increase. To the extent that time-series implications may be drawn from cross-section analysis, concentration may well be a reversible process.²⁶

As expected LOGPOP2 is a good predictor of LOGPOP1. (Other similar models did not result in significant coefficients for added variables representing the populations of either CITY3 or CITY4, which are much more strongly correlated with the population of CITY2—and one another—than with that of CITY1.) In fact, the strong association between the populations of the two largest cities is a confirmation of the commonly accepted notion that their ratio, which measures departure from that association and is often called “primacy,” is an important summary index of urban concentration.

Only 55% of the cities studied are seaports. (Many of the world's largest cities, especially those of Europe, are not ports, a fact which reflects the declining importance of sea trade in urban development.) Nevertheless, the coefficient for this dummy variable implies that port cities are, *ceteris paribus*, 22% larger than inland cities.

On the other hand, 83% of the largest cities *are* national capitals. The CAPITAL variable has been unjustly neglected in previous analyses. There is a powerful (one might almost say mystical) association between primacy (here used in its nontechnical and root sense of being number one) and “capitality.” The link is certainly circular, in that biggest cities are often selected as capitals, while being the capital, as the regression suggests, is likely to stimulate the growth of a city. In this case being a national capital results in a substantial 48% difference in POP1.

It is tempting to speculate about the policy implications of this finding in the light of the slow but steady twentieth century movement to disassociate the capital function from CITY1 (Washington, D.C., is a very early case). The development of what are often called fiat capitals is part of an effort to restrain urban concentration, but obviously there are political motivations for these moves as well. Naturally these results spark the question, Will Nigeria's move of the capital from the coast to the interior reduce the size of Lagos by 48%? Obviously not immediately, especially because Lagos is relatively small (see below); but the effect in the long run, with the growing importance of the public sector in employment, may be significant.

The last variable, METRO, is included because I was able to secure metropolitan population figures only for 58% of the cities. The

coefficient for this variable is barely significant, but it is certainly valuable in the equation to help correct for this problem, and its sign shows that city data tend, as expected, to underestimate the size of the largest cities more than that of the smaller cities.

On the whole, the linear equation is a very satisfying quantitative summary of the factors that determine the population of a nation's largest city. First, the multivariate causal model on which it is based is coherent and reflects much of what we currently know both about why cities grow and about why the largest cities get so big. Second, the explanatory success of the model is demonstrated in a value for *R*-squared, which is quite large given the substantial diversity of the national cases (remember that New York City is 450 times as large as Gaborone, Botswana!). Finally, there is a pleasing balance among the explanatory variables. Each substantive regression coefficient has a (one-tailed) 95% confidence interval which excludes zero, each sign is as expected, and none of the variables particularly dominates the model. The success of the overall result is a reflection not only of logical model building but also of the state of the art of international data gathering.

But what of the 12% of the variance of LOGPOP1 that is not accounted for by the model? The equation leaves a vector of residuals estimating *u* above, given by

$$\text{LOG} \left[\frac{\text{POP1}}{\text{POP1 (predicted)}} \right],$$

which represent the logarithm of the ratio between the actual population of CITY1 and its population as predicted by the model. (See App. for a list of the cities, the residuals, and their computed ratios.) The residuals, which are normally distributed, may be regarded as reflecting our "surprise" at any city's bigness or smallness within its nation, given its characteristics. In this sense the positive residuals are a measure of what is technically called primacy, or the tendency of CITY1 to be larger than expected. (To conform to the Latin origins of the word, I suppose negative residuals might be referred to as examples of "ultimacy," but it is unlikely that this neologism will see much use.) A natural question therefore is, What is the relationship between these residuals and other more traditional measures of primacy? To answer this, I computed six indices of urban concentration that commonly appear in the literature. The definitions of these indices, their mean values, and their correlations with the residuals of the model above are presented in table 5 (the mean of the residuals is of course zero).

There are three points to be noted about these comparisons. First, the correlation of the residuals with the other indices is between .4 and .6, indicating that the residual is as expected a measure of urban con-

TABLE 5

COMPARISONS OF RESIDUALS WITH OTHER MEASURES OF URBAN CONCENTRATION

Index	Mean	Correlation with Residual
POP1/(POP1 + POP2)	.742	.54
POP1/(POP1 + POP2 + POP3)	.647	.62
POP1/(POP1 + POP2 + POP3 + POP4)	.596	.64
POP1/URBAN POP	.386	.46
POP1/TOTAL POP	.149	.40
Rank-size slope*	1.49	.63
Rank-size intercept†	2.93	.95‡

*Slope of the rank-size line $\text{LOGPOP}(k) = a - b\text{LOG}(k)$, estimated for $k = 1, 2, 3, 4$.

†Intercept of the line above.

‡Correlation with predicted value of $\text{LOGPOP}1$.

centration, although not one based only on city population data. But because the model includes so many diverse factors it is clear that the residual, though not orthogonal to the other indices, certainly measures something new about CITY1 in addition to reflecting more commonly accepted measures of urban concentration.²⁷

Second, the residuals seem to reflect a comparison between the sizes of CITY1 and the other largest cities, rather than between CITY1 and CITY2, on the one hand, or the total population, on the other (note the trend of the correlations for the first five indices). Third, the residuals are most strongly correlated with the slope of the rank-size line, which finding indicates that this line, in the absence of data other than city populations, is a useful way to model urban concentration. Note as well that the mean value of the slope is a good deal steeper than 1 and much steeper than the mean value of .88 that Rosen and Resnick derive for their 44 countries.²⁸

The most interesting bivariate characteristic of the residuals, however, is revealed in the pattern that results when the residuals are broken down by world regions. Figure 3 presents the median, quartile, and extreme values for the residuals for five national groupings that correspond to the inhabited continents of the world. Latin America and the Middle East exhibit the highest individual and group levels of the residuals; Africa and Europe (with North America, the USSR, and Oceania) exhibit the lowest; and Asia displays by far the greatest regional variation. Because I have arranged the groupings in order of their mean values of the variable LOGENERGY for each continent, the nonlinear association between this new measure of urban concentration and an index of the level of economic development is revealed. It is hard to resist the temptation of suggesting that Africa is just beginning, while Europe (including the United States, Canada, the USSR, and Oceania) is completing the process of extreme urban concentration of which Asia, the Middle East, and especially Latin America are still in the midst.

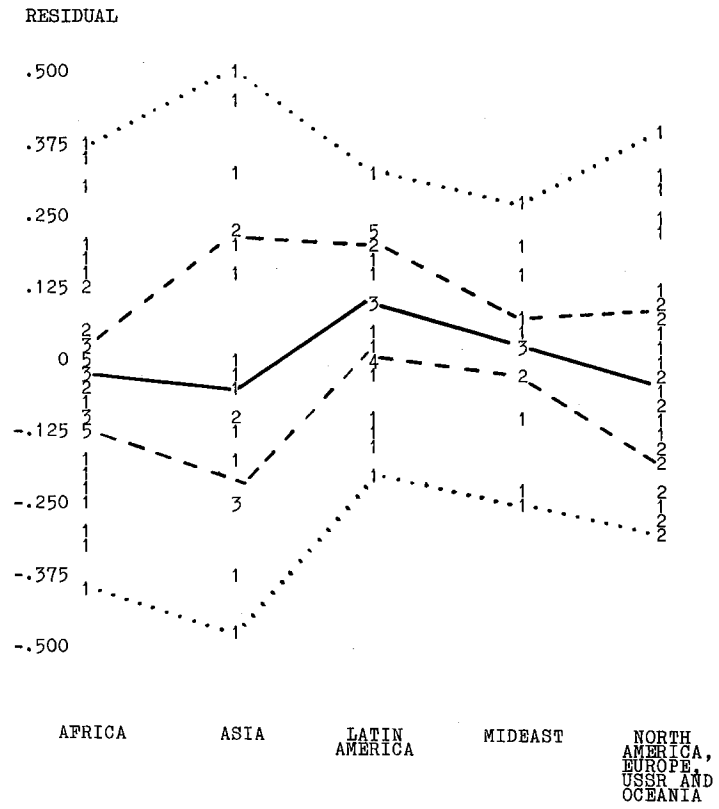


FIG. 3.—Scatter diagram: residual × region. Lines connect extrema, quartiles, and medians.

In order to examine further the nonlinear effect of the development variable, I estimated the coefficients of a second model that included a term for LOGENERGY squared. The results, presented in table 6, show this phenomenon quite clearly. Along the LOGENERGY dimension the predicted values of LOGPOP1 are a parabola with a negative second derivative and a maximum at LOGENERGY = 2.787, close to the sample mean for this index. Furthermore, the estimate for the linear term is much improved, as can be seen by comparing the *t*-statistic from table 4. On the other hand, as we might expect, estimates of some of the other coefficients suffer in this second model. Two of the demographic variables, LOGPOP2 and PCTCHANGE, become less important, as do the spatial variables. I interpret this alternative model as being particularly sensitive to the interaction between development and urban concentration, while the first model is a broader examination of a wider set of spatial, demographic, and political factors.

This finding follows in a long tradition of analyses of spatial inequality and concentration and their relationship to development.²⁹ A

TABLE 6

REGRESSION COEFFICIENTS OF MODEL INCLUDING LOGENERGY SQUARED

Variable	Coefficient	<i>t</i> -Statistic	Contribution to R^2
INTERCEPT	-1.54	-5.44	...
LOGAREA	-.141	-1.71	.0027
LOGLENGTH	.353	1.95	.0035
LOGPOP	.685	10.33	.0988
PCTCHANGE	.0123	.53	.0003
LOGENERGY	.791	4.43	.0182
LOGENERGY ²	-.142	-4.06	.0153
PCTURBAN	.00808	5.76	.0307
LOGPOP2	.0883	1.32	.0016
PORT	.0627	1.73	.0028
CAPITAL	.136	2.59	.0062
METRO	.0743	2.08	.0040

NOTE.— $R^2 = .894$; adjusted $R^2 = .884$.

more definitive explanation of this phenomenon must await further analysis, but two rival hypotheses seem worth exploration. First, there may be severe biases in the data that result in overestimates for POP1 in Asia and Latin America—regions known by all of us to have large cities. Hence overestimates of urban populations may be less discounted here than in other regions. But I would expect that this self-fulfilling prophecy might be avoided in my analysis by the inclusion of POP2. Second, the classical model of a temporary phase of developmental concentration may well hold. Such a process results from the negative feedback (or second-order) effect operating from development via population change and migration (see fig. 1). Economic growth, exogenously determined, leads to early population growth and urban migration, followed later by lower levels of natural increase and by out-migration from the biggest cities as they cease to be prime foci of economic activity.³⁰

Because the power of any model is often revealed by what it does not fully explain, it is useful to highlight the outliers, cities with particularly extreme residuals. (Reference here and elsewhere is to the residuals from the model presented in table 4.) Bangkok is roughly three times its expected size while Phnom Penh is about one-third of its expected size. It is amazing that the extremes are geographic neighbors! Bangkok is notoriously large, having earned the opprobrious epithet "parasitic" in a recent article, for the city is 22 times larger than the combined populations of the next three cities and contains virtually all of Thailand's urban population.³¹ Next door, however, is what remains of Phnom Penh after the Khmer Rouge invaded the city in 1975. The present condition of both cities suggests that political factors are, in the end, very important in shaping the urbanization—or

deurbanization—process. A similar story might be told of North and South Korea: the capital of the latter still swells in the flush of an economic boom, while Pyongyang continues its slow but apparently impressive recovery from its devastation during the Korean War.

What follow are further anecdotal explanations of the position of various outliers in the ranking. Budapest, although in a communist state, is 10 times larger than CITY2. Cairo is a very large city (twelfth in the world, according to my sources) in a continent of relatively small cities. Of Dakar perhaps a repetition of the phrase “Dakar et le désert Sénégalais” will suffice, for Senegal preeminently reflects patterns found in France.³² On the whole, the positive outliers, eight of them over twice their predicted sizes, seem in general to be large cities dominating relatively small countries, but it is interesting that such a diverse assortment of cities is brought together to reveal the nature of this robust index of urban concentration.

Turning to the negative residuals after Kampuchea, it is slightly ironic that this research was done in Nigeria, where Lagos and Ibadan compete for the title of the largest city. To some extent they are twin cities: although their developmental histories are quite different, they are spatially close and functionally complementary in that one is a political and commercial capital while the other is the largest of several ethnic and cultural capitals in Nigeria.³³ However, it is obvious to any visitor to Lagos that the United Nations’ population figure—and most other estimates as well—is drastically too small, as perhaps are most other African urban population figures. If Singapore instead of Kuala Lumpur were counted as CITY1 for Malaysia the country would not be an outlier, which result shows that the model does detect the absence of a certain geographic wholeness in this island nation. Ouagadougou and Niamey are in neighboring countries and are among the smallest capital cities in the world. Moscow is an interesting case: the capital of a vast empire, it would be expected to be very large, but it seems that effective efforts to restrict migration into the city have led to its containment, and of course Leningrad gives the current capital strong competition in many respects.³⁴ Khartoum, to answer a question posed in note 8, is only part of a larger metropolis including the large agglomeration of Khartoum North plus the old city Omdurman. Finally, Rotterdam and Brussels are something of a surprise, but at the scale of the Low Countries, distinctions among the biggest cities seem to be largely statutory.

Although each extreme, whether large or small, may call for its own unique explanation, there is at least one lesson to be learned from the group: political factors can play a significant role in determining the size of a nation’s capital. At the top of a nation’s hierarchy a certain discreteness intervenes resulting in often abrupt changes in function and, consequently, in size. Von Boverter claims that “in the present

situation of the social and environmental sciences the question of maximal city sizes is above all a political decision. . . ."³⁵ Richardson makes this discreteness explicit by observing that "the number of highest order (national and supranational) functions is so limited that there is no scope for more than one or two cities to grow up to the highest rank. In the interurban competition for big size there are few winners because there are so few prizes."³⁶

My attention, for personal and scientific reasons, turns finally to Africa. Although it is the continent that along with Europe has the lowest values of the residuals, Africa is also experiencing several of the developmental conditions that point to high urban concentration in the near future. The pull of the city and the push out of the countryside, coupled with high natural growth rates, present large-city conditions that make the phrase "urban crisis" seem quite inadequate. No other continent, moreover, has so recently felt the hand of colonialism. During the postcolonial period, when the capital cities continue to serve as the umbilicus of the metropolis, it is likely that hyperurbanization at the political and commercial center will continue.³⁷

In this regard it is useful to compare the residuals of the 15 African countries once ruled by Britain with those of the 17 former French colonies. These two groupings constitute over 80% of the African nations in my sample. In table 7 the countries are arranged by colonial language into three classes (according to whether the residuals are greater than .1, between .1 and -.1, and less than -.1). The Francophone countries clearly dominate the large residuals, while the Anglophone countries have an edge among the large negative residuals. (The χ^2 statistic for this table is only 3.44, hardly statistically significant, but this should not distract us from the manifest pattern.)

The differences between the colonial urban policies of France and Britain have been well summarized in a recent article by Gugler and Flanigan, who refer in the case of France to "the hubris of capital cities

TABLE 7

CROSS-TABULATION OF RESIDUALS BY FORMER COLONIAL RULER FOR 32 AFRICAN NATIONS

Residual	Britain	France	Total
Greater than .1	1	5	6
Between .1 and -.1	7	8	15
Less than -.1	7	4	11
Total	15	17	32

NOTE.— χ^2 statistic = 3.44; $p < .20$.

