A Taxonomic Review of Urban Size Distribution Models

by

Lee De Cola

Research Paper 8618

ABSTRACT

The paper uses a consistent notation to organize into a taxonomic scheme over 70 years of theoretical and empirical research into the nature of regional settlement systems. A broad collection of models are classified by the ways in which they treat system structure, developmental processes, and stochastic influences. The scheme reveals the great richness of regional scientific thought and provides a framework within which databases may be organized and future research may be directed.

1. INTRODUCTION

A central concern of regional science is the description and understanding of settlement systems. Much of the literature in this area, whether theoretical or empirical, is usually content to ignore various aspects of the issue of how the system is defined and described, with the result that students of the subject work in a cloud of ambiguity that natural scientists would not accept. What precisely might a writer mean when he refers to one urban system as being "more developed" than another? When writers, from Jefferson (1929) to Bannerjee and Schenk (1984) speak of "primacy" it is often unclear whether they are referring to an attribute of a settlement or of a system. (What, to take a recent example, does Sheppard (1982, p. 135) mean when he says that "the primacy of city i over city (i+1) is greater than the primacy of city (i+1) over city (i+2). . . ."?) And how might we better use the traditional core/periphery distinction of sociology to improve our understanding of migration (Vining 1982) and political dominance (Friedmann 1979)?

This paper uses a consistent notational framework to review four classes of settlement systems models, which may be distinguished by their focus on structure, process, growth, and chance. While it would be possible for the discussion to be completely general, it will help motivate the argument if we try to focus on one facet of the system, namely the distribution of city sizes. This urban characteristic has received more attention than any other during this century, and lively debate even today surrounds the welfare consequences of city size of the micro- and macro-scale. Nevertheless, the discussion could easily be generalized (and perhaps made a bit less absorbing) to a focus on any system characteristic.

We shall beg the question of precisely how the region is designated except to say that we forbid the circularity of an approach that specifies the region on the basis of the settlement definition. The availability of data, in any case, often limits our freedom in this regard (although this constraint is being ameliorated with better remote sensing information). We therefore begin with a spacially connected region (territory, nation, etc.) within whose boundaries human, material, and information flows are maximized and across whose boundaries such flows are minimized. An empirical approach to this problem would be to impose a grid on the region, count the resident population within each cell, and exclude those cells whose population (and hence density) fell below a certain critical value. As this critical value increases (say from 0 to its maximum possible value of $10^7/km^2$) more and more cells would be excluded. But for a range of density values we would find clusters of continuous cells all above the critical value (De Cola 1986). Assume, therefore, that at some density value there remain I clusters, indexed by i.

For each settlement (often called a "city" for variety's sake) let J attributes be measured. These attributes could be measured by variables indexed by j and they might be of the following kinds: 1) binary (0/1) variables indicating whether or not a settlement has some characteristic (such as being a capital), 2) nominal, perhaps describing the city's role or dominant cultural group, 3) ordinal, such as whether the city is a parish or the head of a bishopric, and of course 4) continuous, such as its mean July temperature or its population. The number of such variables would depend on the descriptive and analytical purposes. The variables, as well, might be ordered in a causal hierarchy, but this is not necessary. In addition, for each settlement and for each variable let T measurements be made at equal periods, indexed by t. The magnitude of T, like that of I, would be limited by the range of available records and determined by the periodicity of important events (Dendrinos and Mullally 1981).

Once all these data have been compiled they can be assembled into an IxJxT data structure called X, which may be thought of as a stack of T data matrices, each containing IxJ items of information (Berry 1964 and Hordijk 1979). Although in the discussion that follows the treatment of dimensions may tend to be casual, in general matrices and vectors will be denoted by uppercase letters and scalars by lower case, often subscripted. Consequently, a typical data element would be x_{ijt}. Thus we can interpret y_{it} (= x_{i1t}) as the population of city i at time t (so that $Y = X_{.1.}$). If the data structure is X then a fundamental problem of settlement system theory is to predict Y in terms of: the structure X, a matrix of parameters B, and various error terms U according to the following general formulation

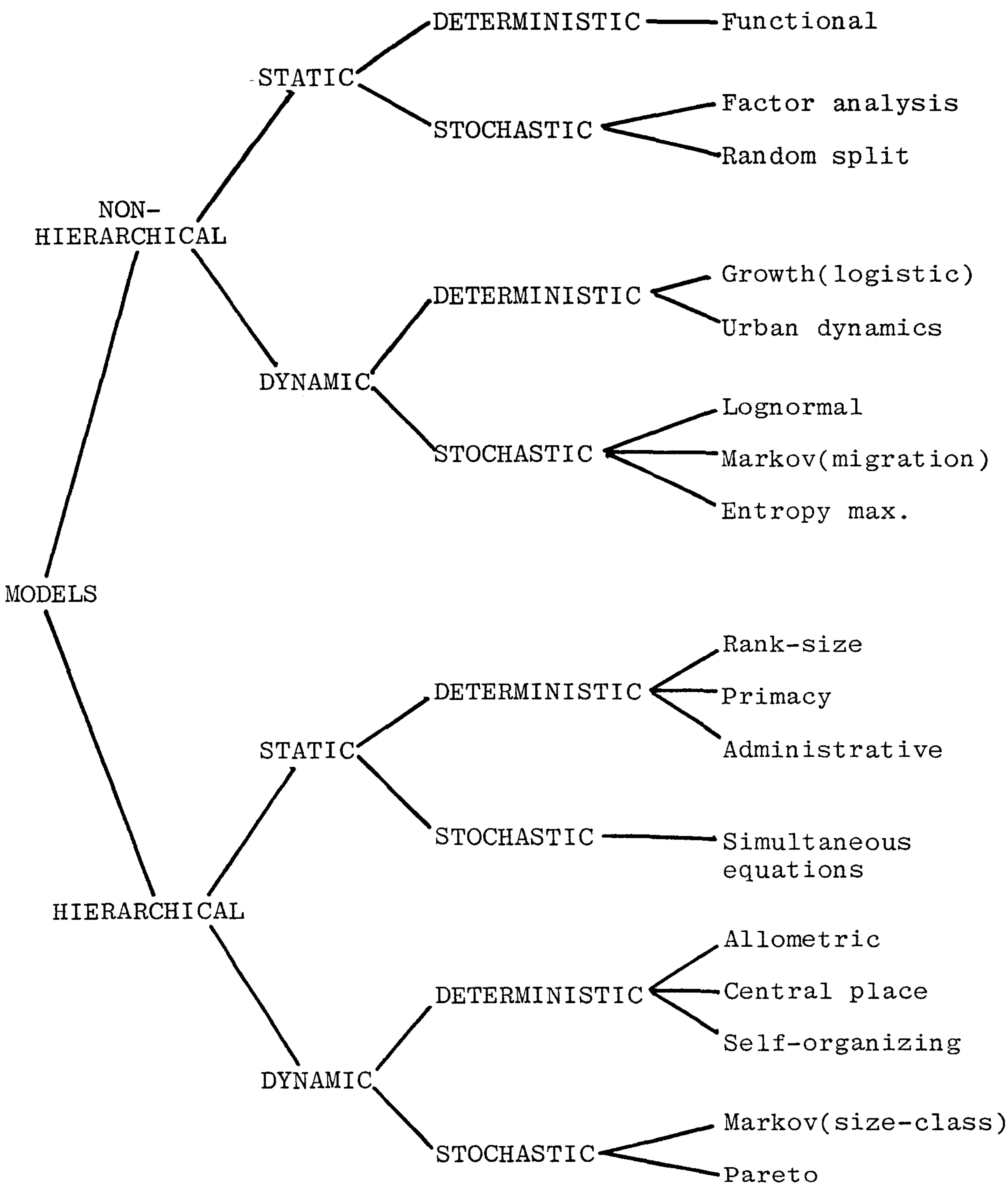
$$Y = f(X,B) + U$$

where X has three dimensions and U is a stochastic term.

2. MODELS OF STRUCTURE AND PROCESS

The organization of X and of the general formulation require that settlement systems models be discussed in terms of their treatment of at least four problems. First, and intimately related to the specification of the system, is the question of whether the analysis is systemic or atomistic: the former models examine the (usually hierarchical) structure of the settlement system, while the latter treat each city in isolation from the others. Second is the matter of dynamic as opposed to static models. A static model might treat only the latest "slice" $t = T$, whereas a dynamic model would deal with changes over time. Third, some models are stochastic, allowing for random influences in system behavior, while others are deterministic. Finally there is the matter of multivariate versus univariate models: is the discussion purely about Y or are relevant values of X allowed to influence the "dependent" variable?

Figure 1 is a diagram which organizes a number of important settlement systems models according to the way they treat structure, process, change, and chance. Obviously there are many other ways of arranging these models, but the table not only captures the notion that model building is a matter of choices but reveals the close



A Taxonomy of Settlement Systems Models

kinship among many supposedly different ways of thinking about the specification of the functional relationship between Y and X. Of course no model deals in detail with all of these aspects of the settlement system--for one thing the data requirements of a complete structure X are linkages is truly overwhelming. But because different models tend to emphasize or the other of these dimensions of the problem, it is possible to organize a broad expanse of regional research in this fashion.

It is important as well to understand the spirit of this taxonomy. Models, unlike organisms, cannot be captured and dissected to reveal their structure; they are intellectual systems. At the present state of our science models tend to reveal as much about the intellectual--and often political--styles of their creators and adherents as they do about settlements and regions. The tree figure is therefore also an expression of style. For example, random split and simultaneous equation models are both regarded as static-stochastic approaches because they focus on the system at a single point in time while acknowledging its random nature. The former model as well is non-hierarchical because no particular ordering or structure is imposed upon the system., while the latter will assert that some elements of the system are more causally important than others. I do not, furthermore, claim evolutionary validity for this taxonomy--although I would of course like to see future model-building, as well as teaching, influenced by this organizational approach.

To conclude this introduction, although the focus of the discussion will be on understanding Y in terms of X, it is worthwhile to begin with the recognition that the population of a city is its most interesting and important characteristic, and that a great deal of attention has been given to the dependence of various other phenomena (economic activity, interaction, congestion, social welfare, etc.) upon Y, according to the general formulation X = f(Y). This is the approach of the city size literature (Alonso 1971 and Richardson 1973). Certainly no analysis of urban systems is complete--or may even begin--without appreciating just how much "size matters."

2.1 Multivariate models But we are more interested in understanding the determinants of y for a given city and of Y for the set of I cities, and the simplest way to begin is to look for a collection of attributes which each city does or does not have. These multivariate attributes may be assembled into an IxJ matrix X of 0/1 bits, such that $x_j = 1$ if city i has function or characteristic j and $x_{ij} = 0$ otherwise. If the characteristics of the cities appear at random among them (i.e. if the probability that a city has a bakery for instance is independent not only of which other cities have bakeries but of what other functions that city has) then the probability $p(x_{ij} = 1) = b$, a constant, and X is a random matrix containing no information in itself. However, the use of the model

$$Y = XB + U \tag{1}$$

to estimate the J coefficients of the column vector B will tell us (as in analysis of variance) the influence of a given characteristic upon urban population, without permitting the reasonable assumption that the size of a city also influences its functions.

Suppose, to constrain things further, that we wish to classify cities into types. If each city is unique in its class then we need I variables, but even $J = \log_2(I)$ binary variables will suffice to distinguish among them. In general however we are likely to find that a few classes have many cities and many classes have a few, for in theory lognormality (as applied to the sizes of classes, not to y) will obtain (Aitchison and Brown 1957, p. 27). Some characteristics j may appear only once (such as being a national capital, although there are exceptions to this rule) so that for each j

$$\sum_i x_{ij} = 1$$

And there are even classical treatments, in what might be called the <u>functional</u> literature (see the tree, and Hudson 1976, Rochefort 1972) which assign a single role to a city: port town, market center, government seat, etc., so that

$$\sum_j x_{ij} = 1$$

for each i.

Although this functional model has been treated somewhat simplistically, we should not ignore the fact that the structure of X is extremely important in understanding the urban system. Covariances among its columns will suggest which characteristics cluster together, perhaps to give hierarchy and certainly to give organization to the system. And scalographic analysis (Abiodun 1967) among the variables will certainly reveal hierarchy among the functions. Even with so simple a binary data matrix a great deal of structure is made evident. If we further assume that people will seek to be close to many functions and that functions will be sited near many people, then we have gone a long way toward understanding why the distribution of city sizes is positively skewed.

If we allow X to contain data about continuous as well as binary variables then the coefficients of model (1) may be estimated by regression to reveal the statistical determinants of city population. This is an analytical approach of tremendous value in increasing our understanding of why some cities are so much bigger than others and how they grew to be so large. But if we are unable to say very much about the causal linkages among the variables, then a particularly fruitful approach is the factor analysis model

$$X = FB + U \qquad (2)$$

where B is an MxJ matrix of factor loadings and M should be considerably smaller than J. Clusters of factor scores may suggest interesting classes of cities, but the dependence of these patterns on the variables chosen should make us wary of mistaking synthesis for discovery. Nevertheless, good use of this technique has been made by Berry (1972) and in West Africa by Mabogunje (1965) and McNulty (1969).

If we are able to impose a sufficient number of causal constraints upon the information to permit the solution of the identification problem of estimating up to J^2

coefficients from J equations, then we can calibrate a simultaneous equation system

$$YB + XG = U$$

where Y may be a matrix (and not merely a vector) of endogenous variables interrelated by B and related through G to a set of exogenous variables X (Johnston 1972). In addition, the variables themselves may be well- or weakly-ordered into a hierarchical structure of stochastic dependencies presumed to reflect dynamic processes, as is done in causal modelling (Moir 1977). By this state we have probably taken the multivariate approach to its present limits, given that very little progress has been made in elaborating such models to explain the behavior of the urban system. Although a simulation does appear to be a successful strategy in the face of insufficient or unreliable data, a bolder econometric use of the data we have would seem fruitful, especially in an effort to improve those data.

The underlying analytical message of this diverse group of multivariate models is that no attempt to analyze urban concentration is complete without a firm understanding of the causal linkages between a city's population and its characteristics (functions, role, etc.). In this sense examining the population of the city itself, and particularly that part of the vector Y that remains unexplained by known factors (De Cola 1984), is an efficient strategy.

2.2 Dynamic models These models assume that systems have memories, and the simplest formulation is a difference equation

$$Y_t - Y_{t-1} = f(X_{t-1}) \qquad (3)$$

so that the change in Y reflects the state of the system in the past. A general discussion of univariate <u>growth</u> <u>models</u> is provided by Boulding (1953). Thus in (3) if f is a premultiplying matrix B we have a linear trend of growth or decline depending on the signs of the elements of the column vector B. In the univariate case, if $y_t - y_{t-1} = b_1 y_{t-1}$ then y grows or declines exponentially according to the analytic expression $y_t = y_0 \exp(b_1 t)$ (Goodman 1974). This exponential or positive feedback model will appear below in the discussion of lognormality. Although this model ignores the fact that no truly closed system can grow, it provides a good rule of thumb approximating the gross rate of natural increase.

A more sophisticated model has $f = b_1(b_0 - y_{t-1})$ in which y is approaching some target b_0 with a rapidity proportional to $b_1 > 0$, so that we obtain <u>logistic</u> behavior. One analytic expression of this model is $y_t = b_0/(1 + \exp(-b_1 t))$ in which the parameters explicitly represent "scale" b_0 and "rate" b_1 (Haggett, Cliff and Frey 1977). Although the teleology of this model is a drawback (what determines b_1?) it does suggest that there are limits to growth in city size. (Incidentally, the derivative dy/dt of this logistic growth model gives the bell-shaped curve which fascinates development scientists.)

We are slightly better off when we allow the magnitude of some system variable x_t to determine y. If for example x is regional population then in the allometric model (Sahel 1978 and Vining and Louw 1978) the growth rate of a city is proportional to that of the region, so that $(dy/y)/(dx/x) = b_1$ and therefore $y_t = b_0(\chi_t)^{b_1}$ (economists will recognize b_1 as elasticity.) The city may be organically regarded as growing slower than, at the same rate as, or faster than x according to whether $b_1 > 1, = 1,$ or < 1. This is a simple notion but with potentially confusing implications if the organic analogy is taken too literally (and what, in any case, determines x?), but it does suggest that the ratio y/x may be worth monitoring as an index of urban concentration within the simple city-region system. A more sophisticated differential equation approach to the problem of city growth in the US will be found in Dendrinos and Mullally (1981).

When the arguments of f are allowed to be a wide range of variables from the structure $X_{..t-1}$, which determines its own development, then $X_t - X_{t-1} = BX_{t-1}$. This system is a simple version of the urban dynamics approach of Forrester (1969). Here there arises quite complex behavior (Bertalanffy 1968, p. 20) that in its "counterintuitiveness" bears stochastic characteristics because of the presence of multifarious feedback loops. Although the general approach has had a volue for a wide range of industrial, urban, and even global systems (Global 2000 1982), it has been criticized for "complexifying" rather trivial structures while offering little analytical insight into convoluted systems (Berlinsky 1976).

The value of such simulation approaches, although their descriptive dimensions are limited and their stochasticity an embarrassment rather than a virtue, is that they make explicit the time dependent causal linkages among subsystems in a heuristically lucid way. While we are rightly skeptical of their forecasts, we must acknowledge the power of such models to make explicit the nature of and interrelationships among mental concepts.

2.3 Systemic models The analytical approaches discussed so far tend to treat each settlement independently or as part of an undifferentiated totality rather than as a member of a system of similar elements. But even the most casual examination of a regional settlement system reveals that not all cities are equal, and in fact the preeminence of various orders of settlements is evidence of a hierarchy. One way to reveal this hierarchy is to sort the cities by population in descending order and to assign to each city i its integer rank number r_i in this list. If we examine the possibility that $y_i = f(r_i)$ then f will obviously be monotonic. The simplest curve-fitting approach to this relationship is

$$y_i = b_0(r_i)^{-b_1} \tag{4}$$

where b_0 is the predicted size of the largest city and b_1 is a concentration index, sometimes called the rank-size slope. This simple model often fits the data quite well, leading to its wide use in theory (Parr 1970), empirical analysis (Zipf 1949--the most notorious example), and even policy (Boal and Johnson 19685).

Certainly no other single urban systems model (with the possible exception of the central place structure) has received more attention than this formulation. Although its origins may be traced (according to Carroll (1982)) to the writings of Auerbach (1913), the elegance of the formulation continues to intrigue analysts in Europe (Lasuen, Lorca, and Oria 1967), Latin America (Vapnarsky 1969), Asia (Berry 1971), the US (Mills 1972), and Africa (Ayeni 1980) to name but a few. Although the circularity of the rank-size model must be admitted, even deeper questions surround the meanings of the parameters and how they relate to the way the regional system is defined. Metaphysical significance has been attached to the value of b_1 in Equation (4) by Zipf and others (Mandelbrot 1983), and although a value close to 1 provides a convenient rule of thumb (often called the rank-size "rule"), more productive theoretical approaches, discussed below, seek to explain probability density in terms of population rather than the other way around.

However, the "scale" parameter b_0 of (4) is the focus of a vast literature, dating at least from the writings of Jefferson (1929) to the present day, concerned with <u>primacy</u>. Let y_{max} be the first, largest city in the ordered list (so that $r_i = 1$). The rank-size model predicts the value of y_{max} as b_0, so that the ratio between the actual and predicted y_{max}/b_0 is one measure of the extent to which this number 1 city surprises us as being larger or smaller than expected. And yet, although such primacy indices, and many other ratios as well, have occupied much attention, at least two issues should be separated. First, it is important and entertaining to explore the reasons why a city may be larger or smaller than empirical regularities predict. But, second, there are good econometric and theoretical reasons for arguing that autocorrelations among urban growth rates make the simple rank-size model inappropriate for purposes beyond the essentially descriptive. These issues are raised by Vining (1976) as stochastic questions that could, but will not, be treated below.

One problem with the literature in this area is that it is often unclear not only what primacy is but also how the system manifests this condition. Barring the (trivial and highly improbable) situation in which two or more cities share the same ranking $r_i = 1$, every system has a maximal city which we may wish to call its "primate" settlement (bearing the same relationship to the other settlements as an archbishop does to his subordinates). There seems no problem with this somewhat poetic terminology: all first cities are tautologically primate. If, moreover, we accept the rank-size formulation, with or without a unity slope, as the norm, then y_{max}/b_0 may be defined as a measure of primacy. I argue that such ratio measures of urban concentration, while interesting in themselves, tell us moderately little substantively about urban systems, although they are excellent diagnostic indicators with clear links to developmental variables at the crossnational level. But it appears that the term primacy has outlived its usefulness in light of our success in developing operational measures of urban concentration to which we may refer explicitly.

The ordering of settlements by population, however, does raise the possibility of asymmetrical interurban relationships, as in the case of an administrative hierarchy. Weber (1958) elaborates a theory that places cities at various levels according to their governmental (or religious, etc.) functions, a concept of obvious importance in federations like the US and Nigeria where there are overlying regional systems of administration (Grabowiecki 1981, Onyemelukwe 1978). In this case the rank number of a city may not be unique to that place, and if these functions attract others according to level then we can expect not only a monotonic relationship between y_i and r_i but also extreme positive skewness in the size distribution as well.

When the hierarchy takes on a specifically spatial structure based on production, distribution, and transportation linkages, we obtain the central place theory elaborated by Losch (1954) and others. (Centrality is added to primacy and capitality!) In this theory each city acquires in the course of its development an integral order number n_i which reflects its functions as well as the number of cities subordinate to it. Empirically we may find that $y_i = b_0(b_1)^{n_i}$ where b_1 is sometimes called a "nesting" factor which reflects the level of urban concentration in the system: $b_1 = 1$ predicts a uniform size distribution of an essentially rural system, and larger values of b_1 predict bigger sizes for the higher-order cities (Beckmann 1958, Alao et al 1977, Beguin and Peeters 1981).

Although central place theory is highly deterministic, as the basis of empirical investigations its predictive success is presumptive evidence of its validity. Less teleological, however, are recent developments in the simulation of self-organizing systems which dynamically generate patterns very much like central place lattices (and biological structures as well) using computer algorithms. Although such models are well developed in the natural sciences (Elder 1976, Haken 1977, Prigogne 1980, Mandelbrot 1983) they are only recently being applied to regional systems (Allen and Sanglier 1979). And throughout this discussion we see once again that population remains the single most important index and cause of hierarchy.

2.4 Stochastic models Up to this point the problem of stochasticity has arisen occasionally in the form of an error matrix U (Equation (1)) into which are swept the residuals from models that obviously cannot predict on target the values of each of the Y (Theil 1971, p. 113). This is a problem for any model whose parameters are estimated from real data. For example, in the factor analytic case (Equation (2)), we choose the number of factors M to be less than J, which in turn is less than I, and we are willing to pay a price, measurable by U'U, for this parsimony. But now we shall examine explicitly stochastic processes in which the true determinants of system behavior are either veiled from view by the grossness of our data or not worth the trouble microscopically to examine.

If we begin with a total regional population N we may divide it in each period into smaller and smaller pieces so that after t periods we have $I = 2^t$ fragments such that

$$\sum_i y_{it} = N$$

(The process is like states creation in Nigeria or any other fixed area federation.) In this random split model we eventually obtain an empirically recognizable size distribution that may bear a strong similarity to much real world data (Cliff and Robson 1978). Aitchison and Brown (1957, p. 26) show that this process, as do so many others, results in lognormality. Although the simulation is conceived of as taking place over time, the model is clearly relevant as a dynamic image of how cities grow.

A Bayesian variant of the random split process arises if we ask what happens if the population N is allowed to occupy I cells at random. The number of ways N individuals may do this is

$$Q = \frac{I!}{\prod_{y=0}^{N} (p(y)y)!}$$

where p(y) is the probability of having a settlement of size y. The maximum likelihood approach of the entropy-maximization model of Curry (1964) and others contains the argument that the most probable distribution of people among settlements is that which occurs under least constrained circumstances. In this regard the treatment is similar to the combinatorial problem of the occupancy of cells by particles (Feller 1950, Sec. II.5); and this likelihood is maximized when

$$\ln(Q) = -\sum_{y=0}^{N} p(y)\ln(p(y))$$

is at a maximum. The maximization of Q yields a Pareto distribution (see below). Although this approach has been criticized by Fano (1969) as ignoring the obviously systemic properties of cities--of course they are more than mere cells--the model has great intuitive appeal and, what is more, leads to a distribution with the necessary shape. As Mandelbrot (1965, p. 332) says, "it is common to find that stochastic models yield results more reasonable than any of their assumptions."

Both the models above are only formally dynamic, but they have little to say about the process of urban growth. If we add a multiplicative stochastic term to the difference equation (3) above then $y_t - y_{t-1} = f(X)u_t$. Furthermore, if $f = 0$ then y remains constant, while if $f = 1$ we have a random walk, which might model the early insecure stages of village development. But if $f = y_{t-1}$ then the change in y is proportionate to its previous value, so that y obeys the "law of proportionate effect" of the <u>lognormal</u> process (Aitchison and Brown 1957). This model implies that the relative change in y is just $(y_t - y_{t-1})/y_{t-1} = u_t$. Now if we let the time interval become infinitesimally short we may integrate

$$\int_{y_0}^{y_T} \frac{dy}{y} = \ln(y_T) - \ln(y_0)$$

so that $\ln(y_T) = \ln(y_0) + u_1 + . . . + u_T$. Finally, by the

Central Limit Theorem (Feller 1950) if the u_t are independently and identically distributed then their sum is normally distributed and we may call the mean of this sum b_0 and its variance b_1^2. The complete model is therefore

$$P(\ln(y) < a) = \int_{-\infty}^{a} \frac{1}{b_1\sqrt{2\pi}} \exp\left(\frac{(x-b_0)^2}{(2b_1)^2}\right) dx$$

The lognormal model has important implications for the study of city size. First, the fact that

$$y_T = y_0 \exp\left(\sum_{t=0}^{T-1} u_t\right)$$

suggests that y is continually increasing, cities may not decline in size. For the study of larger cities in less developed countries this is probably not a severe restriction, but there are enough instances of settlement population decline throughout history and in many regions that the model is clearly not universally appropriate (Vining 1982, Wegener 1982). Second, although the model has two explicit stochastic parameters, the role of y_0 as well is certainly important conceptually and may even be empirically meaningful as a threshold below which growth follows one regime and above which growth reflects proportionate effect. Because the original processes of city foundation are usually shrouded in the mists of time, it is difficult for us to explicate the early stages of the development of a given settlement, but we should be prepared for the existence of this threshold below which a place may not matriculate to urban status.

The lognormal model is used by Berry (1961) in his seminal study of size distributions and development, and more recently by Parr and Suzuki (1973) in their work on US city size. It is a parsimonious way not only to capture stochastic behavior but also to summarize the data parametrically, as is done in De Cola (1985) for macroregional distributions of city sizes.

A problem with the lognormal model, however, is that it allows settlements to grow but not to be born. In an evolutionary sense the regional population may grow as space becomes filled with newborn small settlements, but when the territory is virtually completely occupied by urban fields then population will begin to reflect the positive feedback effects of lognormality (Allan and Starr 1982). This evolution was treated theoretically by Simon (1955) and although the resultant distribution has many names depending upon the precise formulation, the term Pareto process is probably most common. The distributional expression of the process utilizes the rank-size regularity but makes population a predictor of rank:

$$r_i = b_0((Y_i)b_1$$

where $b_1 < 0$ is a concentration index such that larger values of b_1 connote greater levels of population concentration in the largest cities (compare to (4)). Recall that I is the total number of cities. Now if we let $n_i = I + 1 - r_i$ be the order of a city (its population on an ascending list of populations) and let

$$p(Y) = \frac{n_i}{I}$$

be the cumulative distribution of y, then the probability density function is $p(y) = (-b_0 b_1 (y)^{b_1 - 1})/I$. Quandt (1964) observes that this formulation could be called the Pareto distribution "of the first kind," there being two other versions involving a third and fourth parameter, in the manner of the elaborations of the lognormal distribution.

In contrast to many of the models discussed so far, the Pareto process is explicitly dynamic in the sense that the shape of the frequency distribution may evolve over time in different ways depending upon the relative influences of city birth and growth. If the former process dominates, then the size distribution tends to be clustered with many small settlements (as in an early territorial phase), whereas if growth dominates we tend to get a few very large cities, and in the limiting case lognormality obtains (Parr and Suzuki 1973). Unlike the rank-size formulation, the Pareto approach avoids the circularity of predicting size from rank, which is itself a function of size, etc. Conceptually, the model explains that big cities are interesting not only because of their size but also because they are so rare, being the uniquely most "successful" competitors in the population race. Furthermore, there is no reason to suppose a discontinuity in the growth process as we ascend the size hierarchy, for as Mandelbrot (1965, p. 332) remarks: ". . . one need not consider _a priori_ that the largest and most important phenomena are different _in nature_ from the small 'noise'."

Certainly the most comprehensive recent empirical treatment of the Pareto process is that of Rosen and Resnick

(1980), who estimate b_1 for the largest 50 cities of each of 44 countries. Bussière and Stoval (1981), in an otherwise somewhat contrived derivation of the Pareto distribution from the Weber-Fechner law of perception, suggest that a city's population "qualifies" it for a certain rank; this notion as well places r_i on the lefthand side.

A final stochastic model, which permits individual cities to retain their identities as the loci of births and deaths and as migration origins and destinations, is the Markov process of interregional flows. If Y_t is an I-element vector of populations and B is a square matrix of migration probabilities, then we may estimate $Y_t = BY_{t-1}$ (Rogers 1978). It should be noted that the random aspect of this process lies in our not knowing precisely who will move where or when, but the growth regime of each city is quite determinate. In addition, because Markov models are based on the examination of micro behavior of migrants, their data requirements are substantial: 1000 cities for example are linked by almost 1,000,000 flows. A more hierarchical approach (the stochastic version of central place theory) is used by Lever (1973) and Robson (1978) in their treatments of the movement of cities between ordered size classes. Okabe (1982) presents a concise discussion of the formalities of this process.

3. CONCLUSION

This attempt to organize some of the regional science research of the 20th century is actually an exercise in "metamodelling" rather than modelling itself. The use of a taxonomic tree and the way important models have been hung on it has served several useful purposes. First, the process clearly deserves to be carried further, and I welcome not only efforts to reorganize the structure but also to place existing efforts within the scheme. As a relational database management structure the scheme could here serves as a taxonomic skeleton. But more substantively, the three dimensions of the data structure X help to organize our thinking about any regional science phenomenon. The I dimension entails specifying exactly what the system is: what does a region include and exclude, what precisely shall be denoted a settlement, and how do varying specifications of each of these elements modify our thinking about the others?

As tentative decisions about these issues are being made, we may go on to consider the J dimension listing the variables to be measured. Data availability constraints are continually becoming more slack while research techniques multiply, so that the need for caution becomes more urgent. We must be particularly cautious about violations of analytical assumptions, misspecified models, and overelaborate description. But the taxonomy helps to sharpen our sense of historical development in this research realm and so will enable us to place our endeavors within a broadly shared scheme of discourse.

The metamodelling scheme obviously needs improvement, and the areas of need point to obscure regions in the continent of our thought. We urgently need better ways of organizing models within the stochastic/dynamic realm. Multivariate models tend to focus on cross-sectional data estimating static relationships for a particular time-slice; while dynamic models usually examine a single system over time using a few variables. A simple example of this dichotomy is seen above in the expositions of growth and lognormal models. More theoretical work in this area should make us more comfortable with dynamic extrapolations of cross-section results and extra-regional conclusions drawn from small scale time series analysis. And of course more empirical work will always improve our conceptualizing.

Finally, the taxonomy has important implications for future research. The data structure X suggests convenient ways of organizing multivariate, spatial, and temporal information for the analytical calibration of a wide range of models. (The near universality of microcomputers gives us instantaneous accessibility to one another's databases: standardization of formats will be especially valuable to regional scientists in less developed countries.) I am sure that almost all researchers, including most of those whose work is cited here, have had to reinvent the wheel by organizing someone else's data for their own purposes rather than by tapping into an existing database. A logical next step after this synthesis would be to provide the means for a donor database to receive prediction and residual matrices from recipient analyses, so that we might extract further variance from one another's results.

More substantively, efforts today to fill out the taxonomic tree should show us which research areas are receiving how much attention as well as how research has shed light on various formulations of the settlement system problem over time. Obviously our goal is models that examine inherently hierarchical phenomena with increasing degrees of spatial and temporal resolution and decreasing stochastic components. The tree is a map to this goal, which recedes only slightly more slowly than the rate of our intellectual progress.

This paper has attempted to shed light on these issues, not so much by proposing new schemes or definitions but by reviewing key ideas of the regional science literature according to an inductive taxonomy that organizes regional settlement systems models according to how they treat the structure of the system and the process of its evolution. The necessity for such a discussion is a manifestation of the relative disorder of regional science at the moment. More to the point, the number of leaves on the branches of this taxonomic tree bear witness to the richness of our discipline.

REFERENCES

Abiodun, J.O. 1967. Urban hierarchy in a developing country. Economic Geography 43:347-367.

Aitchison, John and J.A.C. Brown 1957. The lognormal distribution. Cambridge: Cambridge University.

Alao, N. et al 1977. Christaller central place structures: an introductory statement. Evanston: Northwestern University Studies in Geography, No. 22.

Allen, P.M. and M. Sanglier 1978. Dynamic models of urban growth, I and II. Journal of Social and Biological Structures 1:265-280, 2:269-278.

Allen, T.F.H. and Thomas B. Starr 1982. Hierarchy: perspectives for ecological complexity. Chicago: University of Chicago.

Alonso, William 1971. The economics of urban size. Papers and Proceedings of the Regional Science Association 26:67-83.

Auerbach, F. 1913. Das gesetzde bevolkungskonzentration. Petermann's Geographiche Mittlelungen (Gotha) 59:73-76.

Ayeni, "Bola 1980. Central place theory, rank-size relationships and urban systems in Nigeria. Nigerian Geographical Journal (Ibadan) 23(1,2):127-146.

Banerjee, T. and S. Schenk 1984. Lower order cities and national urban policies: China and India. Environment and Planning A, 16: 487-512.

Beckmann, Martin 1958. City hierarchies and the distribution of city size. Economic Development and Cultural Change 6(3):243-248.

Beguin, Hubert and Dominique Peeters 1981. Urbanization in some hierarchical models. Regional Science and Urban Economics 11:19-37.

Berlinsky, David 1976. On systems analysis. Cambridge: MIT.

Berry, Brian J.L. 1961. City size distributions and economic development. Economic Development and Cultural Change 9:573-588.

_____ 1964. Cities as systems within systems of cities. In Regional development and planning, eds. John Friedmann and William Alonso, pp. 118-137. Cambridge: MIT.

_____ 1971. City size and economic development: conceptual synthesis and policy problems, with special reference to South and Southeast Asia. In Urbanization and National Development, eds., L. Jacobson and V. Prakash. Beverly Hills: Sage, pp. 111-155.

_____ 1972. Latent structure of urban systems. In _Systems of Cities_, eds., L.S. Bourne and J.W. Simmons, pp. 220-232. New York: Oxford University.

Bertalanffy, Ludwig von 1968. _General systems theory_. New York: Braziller.

Boal, F.W. and D.B. Johnson 1965. The rank-size curve: a diagnostic tool? _Professional Geographer_ 17(5):21-23.

Boulding, Kenneth E. 1953. Toward A general theory of growth. _Canadian Journal of Economic and Political Science_ 19:326-340.

Bourne, L.S. and J.W. Simmons 1978. _Systems of cities_. New York: Oxford University.

Bussiere, R. and T. Stoval 1981. A city size distribution model based on the Weber-Fechner law of marginal effects. _Environment and Planning A_ 13:689-701.

Carroll, G.R. 1980. National city size distributions: what do we know after 67 years of research? _Progress in Human Geography_ 6(1):1-43.

Cliff, A.D. and B.T. Robson 1978. Changes in the size distribution of settIements in England and Wales. _Environment and Planning A_ 10:163-171.

Curry, L. 1964. The random spatial economyy an exploration in settlement theory. _Annals of the Association of American Geographers_ 54:138-146.

De Cola, L. 1984. Statistical determinants of the population of a nation's largest city. _Economic Development and Cultural Change_ 33(1).

_____ 1985. Lognormal estimates of macroregional city size distributions, 1950-1950. _Environment and Planning A_, 17: 1637-1652.

_____ 1986. Fractal simulation of urbanized regions. Department of Geography, University of Vermont.

Dendrinos, D.S. and H. Mullally 1981. Evolutionary patterns of urban population. _Geographical Analysis_ 13:328-344.

Elder, J. 1976. _The bowels of the earth_. London: Oxford University.

Fano, P. 1969. Organization, city size distributions, and central places. _Papers and Proceedings of the Regional Science Association_ 22: 29-38.

Feller, William 1950. _An introduction to probability theory and its applications_, Vol. I. New York: Wiley.

Forrester, Jay 1969. _Urban Dynamics_. Cambridge: MIT.

Friedmann, John W. On the contradictions between city and countryside. In Hendrik Folmer and Jan Oosterhaven, eds., Spatial Inequalities and Regional Development. Boston: Martinus Nijhoff.

_____ and W. Alonso 1964. Regional development and planning. Cambridge: MIT.

Global 2000 Report to the President 1982. Harmondsworth, England: Penguin.

Goodman, M.R. 1974. Study notes in system dynamics. Cambridge, MA: Wright-Allen.

Grabowiecki, R. 1981. On the physical development plan for Oyo State. Ibadan, Nigeria: UNDP/Ministry of Lands and Housing, mimeo.

Haggett, P., A.D. Cliff, and A. Frey 1977. Locational analysis in human geography. London: Edward Arnold.

Haken, H. 1978. Synergetics. Berlin: Springer-Verlag.

Hordijk, L. 1979. Problems in estimating econometric relations in space. Papers and Proceedings of the Regional Science Association 42:99-118.

Hudson, F.S. 1976. Geography of settlements. Estover, England: MacDonald and Evans.

Jefferson, M. 1929. The law of the primate city. Geographical Review 29:226-232.

Johnson, G.A. 1980. Rank-size convexity and system integration: a view from archaeology. Economic Geography 56(3):234-247.

Johnston, J. 1972. Econmetric methods, 2nd edition. New York: McGraw Hill.

Lasuen, J.R., A. Lorca and J. Oria 1967. City size distribution and economic development. Ekistics 24:221-226.

Lever, W.F. 1973. A Markov approach to the optimal size of cities in England and Wales. Urban Studies 10:353-365.

Losch, A. 1954. The economics of location. New Haven: Yale.

Mabogunje, A.L. 1968. Urbanization in Nigeria: a constraint on economic development. Economic Development and Cultural Change 13(4):412-438.

Mandelbrot, B. 1965. A class of long-tailed probability distributiods and the empirical distribution of city size, in Mathematical explorations in behavioral science, eds., F. Massarik and P. Ratoosh, pp. 322-332. Homewood, IL: R.D. Irwin.

_____ 1983. The fractal geometry of nature. New York: W.H. Freeman.

McNulty, M.L. 1969. Urban structure and development: the urban system of Ghana. Journal of Developing Areas 3:159-176.

Mills, E. 1972. Urban economics. Glenview, IL: Scott-Foresman.

Moir, H. 1977. Dynamic relationships between labor force structure, urbanization, and development. Economic Development and Cultural Change 26(1):25-41.

Okabe, A. 1982. Toward a more explicit stochastic model of a city-size distribution: a reply to Kim's critique. Environment and Planning A 14(8):1121-1124.

Onyemelukwe, J.O.C. 1978. Urbanization in a development context: patterns, problems, and prospects, in Urbanization and Nigerian Economic Development, Ibadan: Nigerian Economic Society.

Parr, J.B. 1970. Models of city size in an urban system. Papers and Proceedings of the Regional Science Association 25:221-253.

_____ and T. Suzuki 1973. Settlement distributions and the lognormal distribution. Urban Studies 10:335-352.

Prigogine, I. 1980. Being and becoming. San Francisco: W.H. Freeman.

Quandt, R.E. 1964. Statistical discrimination among alternative hypotheses and some economic regularities. Journal of Regional Science 5(2):1-23.

Richardson, H.W. 1973. The econmics of urban size. Westmead, England: D.C. Heath.

Robson, B.T. 1978. The growth of cities in England and Wales. In Systems of Cities, eds., L.S. Bourne and J.W. Simmons, pp. 126-149 New York: Oxford University.

Rochefort, M. 1972. Typologie functionelle des villes en Afrique noire et Madagascar. In Centre Nationale de la Recherche Scientifique, La croissance urbaine en Afrique noire et a Madagascar. Paris: CNRS.

Rosen, K. and M Resnick 1980. The size distribution of cities, the Pareto law and primate size. Journal of Urban Economics 8:165-186.

Sahal, D. 1978. A formulation of the Pareto distribution. Environment and Planning A 10(12):1363-1376.

Sheppard, Eric 1982. City size distributions and spatial economic change. International Regional Science Review, 7(2): 127-151.

Simon, Herbert A. 1955. On a class of skew distribution functions. Biometrika 42:425-440.

Theil, H. 1971. Principles of econometrics. New York: Wiley.

Vapnarsky, C.A. 1969. On rank-size distributions of cities: an ecological approach. Economic Development and Cultural Change 17:584-595.

Vining, D.R. 1976. Auto-correlated growth rates and the Pareto law: a further analysis. Journal of Political Econmy 84(2):369-380.

_____ 1982. Migration between the core and the periphery. Scientific American 247(6):36-45.

_____ and S.J.H. Louw 1978. A cautionary note on the use of the allomctric function to estimate urban populations. Professional Geographer 30(4):365-370.

Weber, Max 1958. The city, translated and edited by D. Martendale and G. Newirth. New York: Free Press.

Wegener, M. 1982. Modeling urban decline: a multilevel economic-demographic model for the Dortmund region. International Regional Science Review 7(2):217-241.

Zipf, G.K. 1949. Human behavior and the principle of least effort. Cambridge: Addison-Wesley.