TOWARD A THEORY OF POPULATION DISTRIBUTION

Arnold Faden

Labor Research Seminar

Prof Gary S Becker

3 December 1959

TOWARD A THEORY OF POPULATION DISTRIBUTION

In the following, only aggregate populations and population densities are treated, no variations by age, sex, ethnic group or economic status being considered. Only United States data are used, but we hope that the results are of more general applicability. The tentative nature of all conclusions reached will be only too obvious.

The best known generalization about population distribution is the Rank-Size law, which states that if the cities of a region are ranked in decreasing population order--the largest being #1, the next largest #2, etc--then the product of a city's rank and its size is a constant for all cities, at any moment of time. This rule was observed as early as 1913 by Auerbach, later by Lotka, Singer and Gibrat, and stressed (or made notorious) by Zipf (National Unity and Disunity; Human Behavior and the Principle of Least Effort). As an illustration, consider the distribution of city sizes for the United States in 1950:

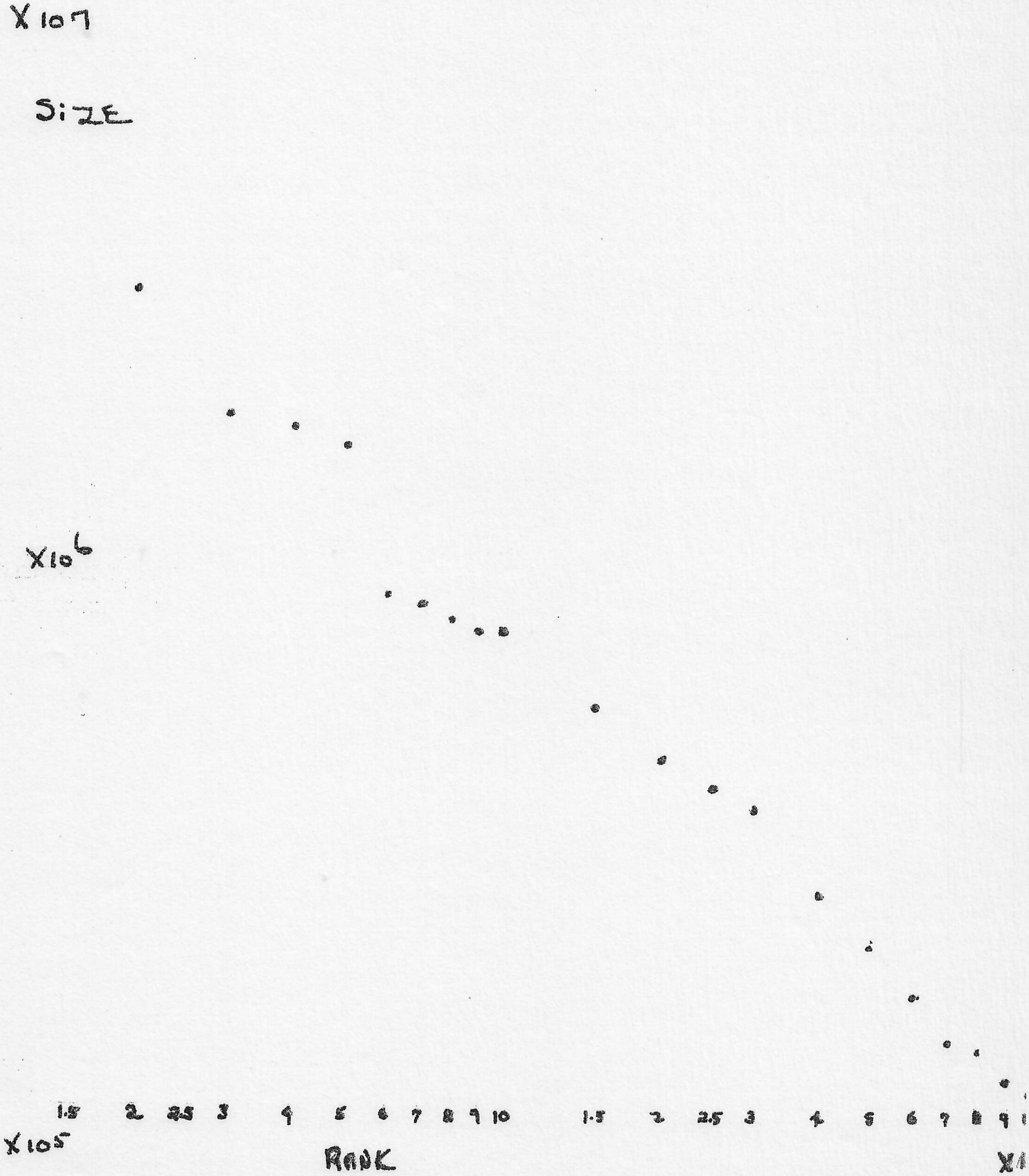
Size Category	Frequency	Cumulated Freq.	Lower limit x CF
1000000 plus	5	5	5000000
500000-1000000	13	18	9000000
250000-500000	23	41	10250000
100000-250000	65	106	10600000
50000-100000	126	232	11600000
25000-50000	252	484	12100000
10000-25000	778	1262	12620000
5000-10000	1176	2438	12195000
2500-5000	1846	4284	10710000
1000-2500*	4615	8899	8899000

*including urban places according to new definition

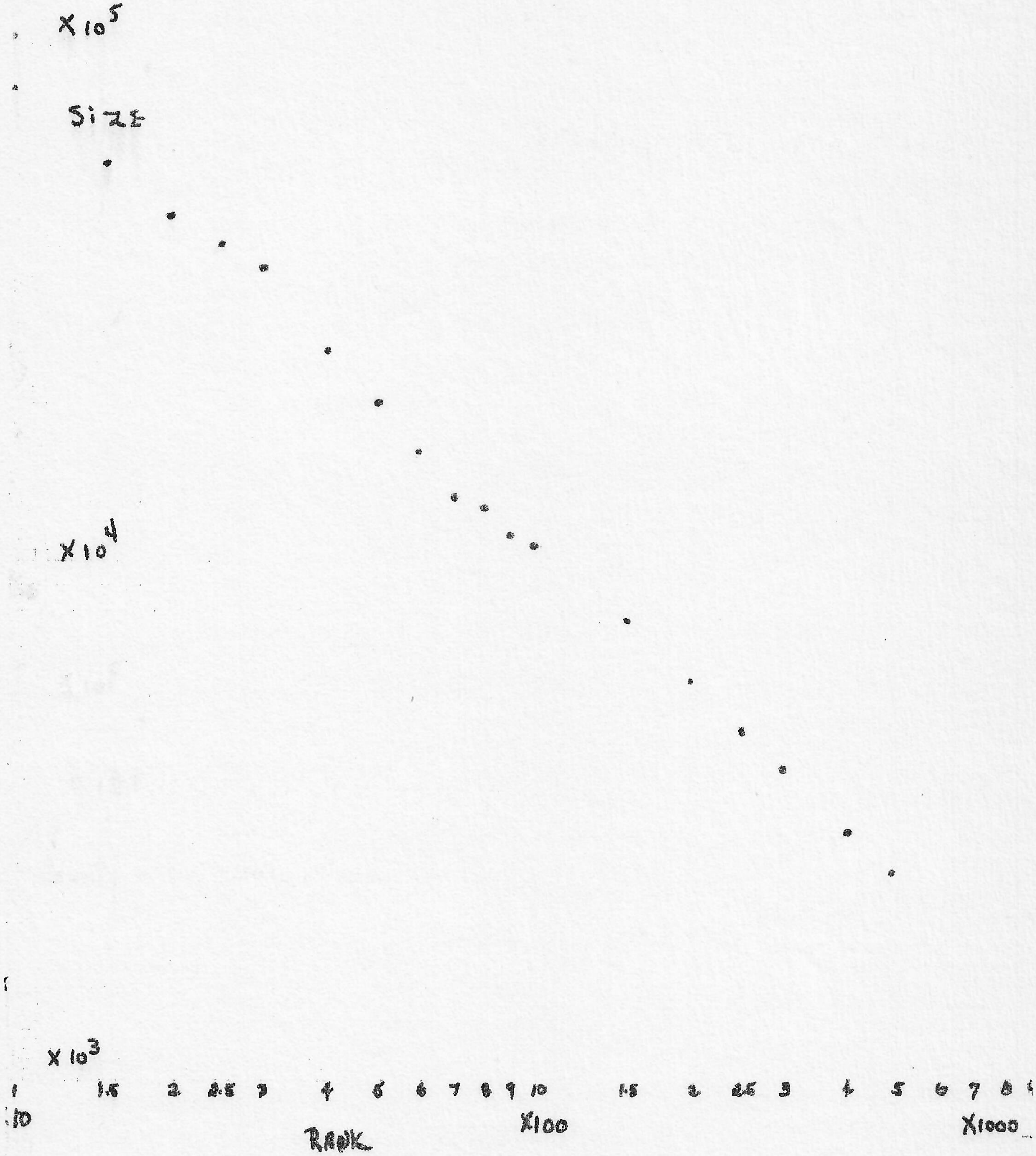
To apply the Rank-Size law to these figures, we assume that the smallest city in any size category actually had the population of the lower limit of that category. The error so introduced is clearly negligible except perhaps for the 1000000 plus category. The cumulated frequency down to and including a category is the rank of the smallest city in that category. Therefore, according to the law, the product of cumulated frequency and lower limit should be a constant: the numbers in the last column should all be the same. Clearly they are not; but they are *almost* the same for a broad middle stretch of cities, say from sizes 5000 to 100000. There is, furthermore, a characteristic tailing off of the product at both the upper and lower ends of the range--a feature which is almost universal in distributions of this kind. The same general pattern, with approximate constancy in the middle range, holds in fact for all census years from 1790 on.

It is tempting to explain the deviations from the rule by pointing out that only the politically defined city is enumerated as such in the Census. This procedure excludes the dense urban fringes of the larger cities and so understates the population of the "natural" city to which, presumably, the Rank-Size rule applies. Similarly, at the lower end there is a possible systematic under-enumeration by excluding people in the surrounding countryside who are directly dependent on the town. Seasonal shifts in population may be pertinent. None of this, however, has been tested.

The question arises as to whether spatial entities other than cities follow a Rank-Size rule. We have graphed the distributions for Standard Metropolitan Areas, for counties and for states, all in 1950. SMA's conform rather well, and this time all the way to the top, which fact supports the urban fringe hypothesis stated above; there is again a characteristic tailing off at the lower end. States are less regular, but show approximate linearity on double-log paper down to state #31, after which they tail off drastically. Counties show a marked linearity down to about county #2000 (that is, about 2/3rds of the way down, as with states), after which they fall precipitously.



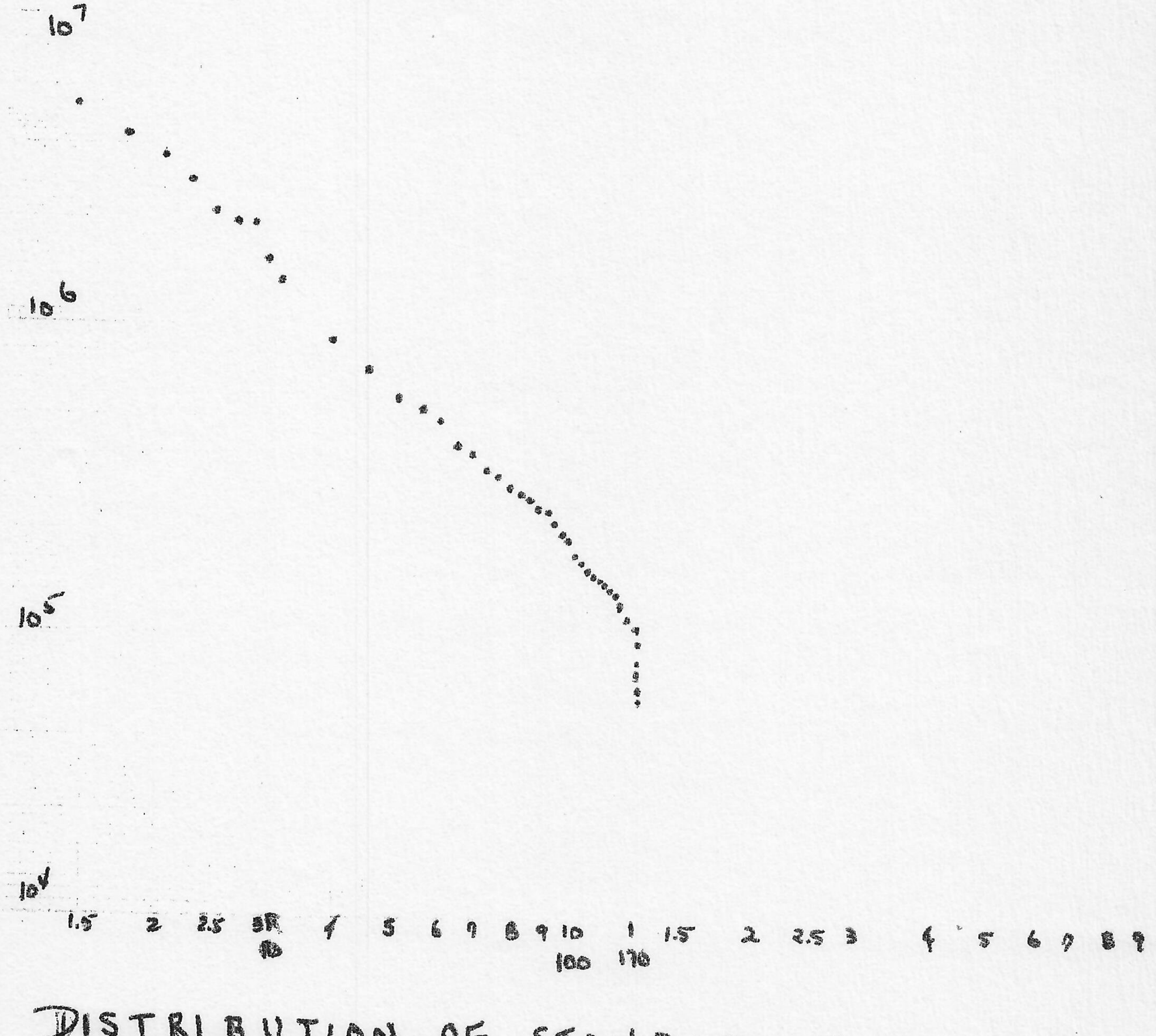
DISTRIBUTION OF CITY SIZES, 1950, FOR RANKS 1-100



DISTRIBUTION OF CITY SIZES, 1950, FOR RANKS 10 - 1000

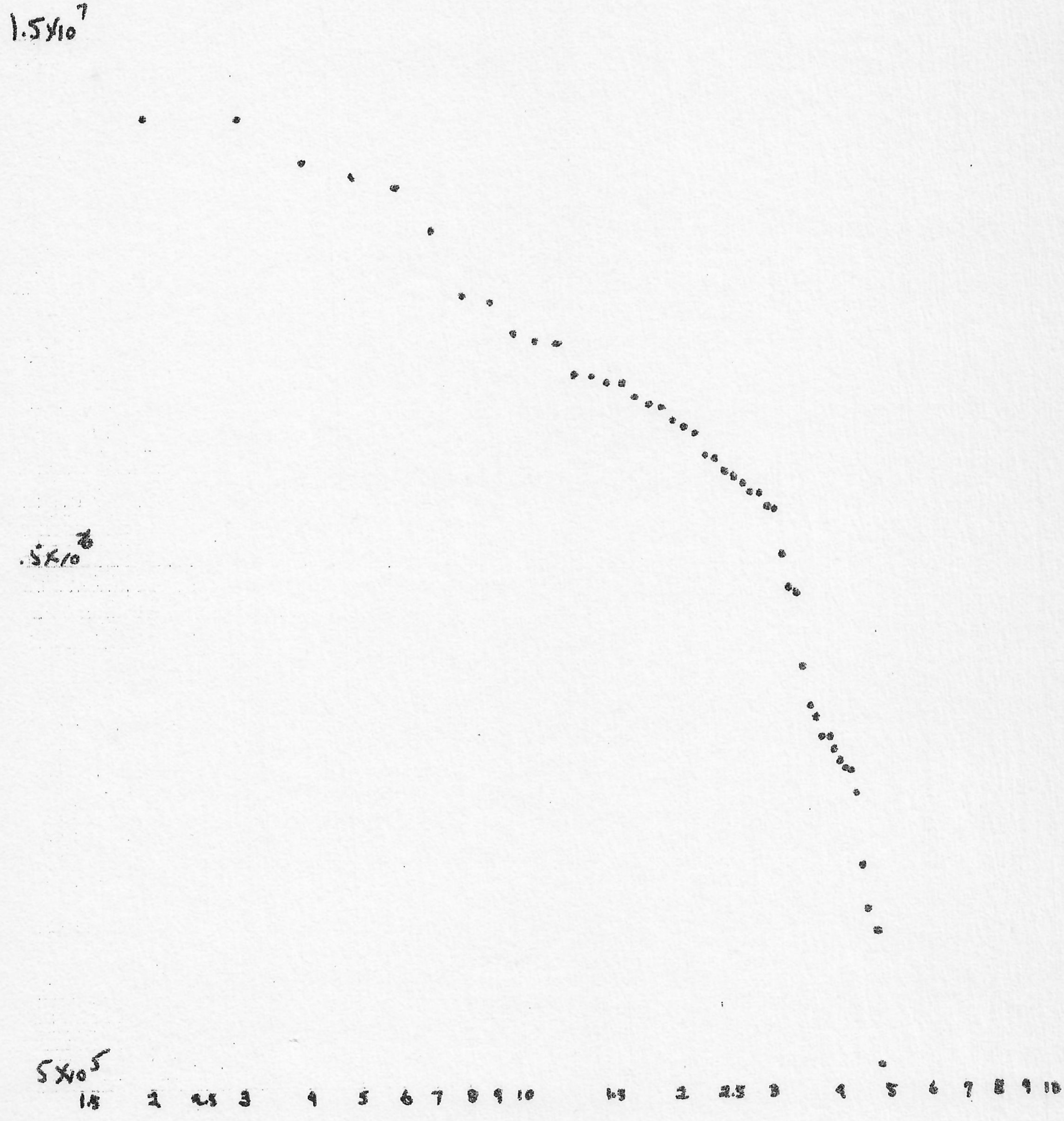
-.95

County + City Data Book, 1952 Appendix A



DISTRIBUTION OF STANDARD METROPOLITAN AREA SIZES, 1950

-.93



DISTRIBUTION OF STATE SIZES, 1950

The slopes in the latter three cases are -.93 for SMA's, -.82 for counties, and -.67 for states. The deviations from the ideal value of -1, which is indicated by the Rank-Size law, can be explained in part, and perhaps completely, by the negative correlation between the population densities of these places and their areas; this fact makes the populations of these places fall off more slowly than their densities. There is reason to believe that densities follow the Rank-Size law (see below), so that populations will fall off with an absolute slope of less than -1, as observed. But a full verification of this would require a theory of areas, which we do not yet possess.

Supposing for a moment the validity of the Rank-Size rule, we may next concern ourselves with the behavior of the rank-size constant over time (the constant being the rank-size product for any city). Since the rule is not satisfied exactly, however, it is necessary to take some typical value to represent a point in time. We select the value for P equals 5000, firstly because a low P enables us to carry the results back to 1790 without strain, and secondly because the value for 5000 has always been in the broad middle range for which the rule holds rather well. The results are as follows, together with total population :

Census year	1790	1800	1810	1820	1830	1840	1850	1860
Population (000000)	3.93	5.31	7.24	9.64	12.9	17.1	23.2	31.4
R-S Constant (00000)	0.60	1.05	1.40	1.75	2.80	4.25	7.35	11.4

	1870	1880	1890	1900	1910	1920	1930	1940	1950
Pop	38.6	50.2	63.0	76.0	92.0	106	123	132	151
R-S	17.7	23.6	35.0	45.0	60.2	73.4	92.0	102	123

If we now calculate $k^3\#^{-2}$, where k is total population and # is the R-S constant, we obtain:

Census year	1790	1800	1810	1820	1830	1840	1850	1860
$k^3\#^{-2}$ (000000000)	16.9	13.6	19.4	29.3	27.2	27.6	23.2	23.8

	1870	1880	1890	1900	1910	1920	1930	1940	1950
$k^3\#^{-2}$	18.3	22.7	20.4	21.7	21.5	21.8	21.9	21.8	21.7

After jumping around a bit in the early years, the product begins to settle down around 1850, and becomes remarkably stable in the 20th century. The closeness may be expressed as follows, if we assume the product was as stable between census years as in them: If at any time in the 20th century someone had given us the population of the country, we could have predicted #, hence the number of cities with more than 5000 people, to within ½ of 1%! Can this be a coincidence? The situation is peculiar because we have as yet no explanation whatsoever for this three-halves regularity.

We can work the other end of the stick by making the assumption that we have here a genuine law for "normal" conditions and then correlating past deviations from it with historical situations. What the product $k^3\#^{-2}$ refers to is of course the gradual urbanization of the country as it becomes filled up with people. The "abnormal" situation of the 19th century would be the great westward expansion of the country which begins on a mass scale after 1800. Instead of filling up the cities people would move into unoccupied territory, which would hold down <u>#</u> relative to <u>k</u> and move $k^3\#^{-2}$ above its "normal" value of about 21.7 billion people. The table would indicate that the greatest percentage westward expansion occurred in the decade 1810-1820. The movement was retarded in the Civil War decade 1860-1870. Finally, "abnormal" conditions came to an end with the closing of the frontier around 1890, which accounts for the stability of the product since then. But of course this is all speculation.

So much for the data. They are intriguing and somewhat mysterious, and call urgently for an explanation. Ideally one would like to have a theory from which the previous results and others could be derived from the maximizing behavior of the multitudes of individual actors who make up the population. Unfortunately such a theory does not exist at the moment. Between this ideal and our data, however, there can be constructed several intermediate levels of explanation--in particular, explanation by migratory propensities. This is the realm with which the remainder of this paper will concern itself. But first we must take a closer look at the Rank-Size and related laws.

The Rank-Size law may be written

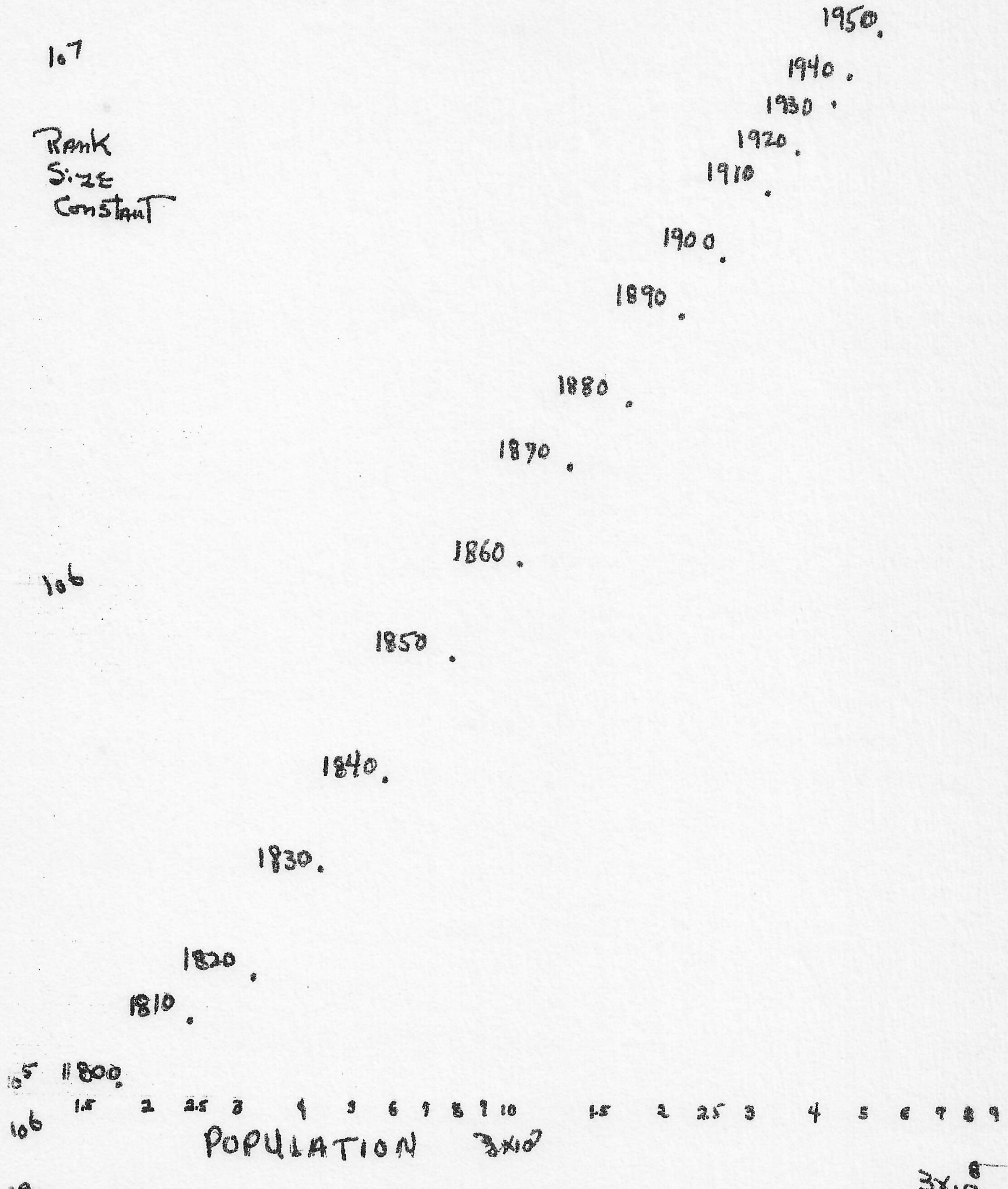
(1) RP_R^n equals #,

where R is the rank, P_R is the population of the Rth ranking city, # is the R-S constant, and the exponent n equals 1. By substituting other values for n we obtain a related family of laws, namely, those which map linearly on a log-log graph. Thus sections of our SMA, state, and county distributions may be described by a law of type (1) with n not equal to 1. (In fact, $-1/n$ is the slope of such a graph, with R plotted on the horizontal axis.)

Equations (1) are, in fact, identical with a family well known to economists, namely, the family of Pareto distributions. This becomes obvious if (1) is rewritten as

(2) R equals $\#P_R^{-n}$,

for the rank R is the same as the number of cities (or income-recipients) with populations (or incomes) greater than or equal to the population (or income) of the city (or income-recipient) in question. The general class of distributions (1) or (2) will be referred to as Pareto distributions, as opposed to the specific R-S distribution with n equalling 1.



190.
475

RANK SIZE CONSTANT AGAINST POPULATION FOR U.S. CENSUS YEARS

Theorem 1: Suppose we have two regions with cities (or other entities) distributed according to a Pareto law, both with the same exponent n. Now suppose we consider the two regions as one composite region, which means we splice the two rankings of cities together and rerank them accordingly. Then the resulting composite distribution will, with small error, again be a Pareto distribution with the same exponent n, and with the composite constant # being the sum of the two constants $\#_1$ and $\#_2$.

Proof: Taking our law in form (2), a city of a certain size P will be exceeded by about $\#_1 P^{-n}$ cities in the first distribution and by about $\#_2 P^{-n}$ cities in the second distribution, hence by about $(\#_1 \text{ plus } \#_2)P^{-n}$ in all. QED

(These proofs are not exactly rigorous, but instructive nonetheless). The theorem extends immediately to a composition of any number of separate Pareto's, providing they all have the same exponent n. In particular, the composition of any number of R-S distributions gives again an R-S distribution.

This persistence of the Pareto laws is one of the great secrets of their ubiquity. As an application, suppose we have found that cities of a certain region obey the R-S law. Then if they are distributed randomly over the region, a coarser meshing of subdivisions would again have the R-S distribution, providing about the same number of cities are caught in each region of the mesh. This goes part of the way to explaining the rough persistence of the R-S distribution among SMA's, counties and states--which distributions are, however, distorted by other factors, notably by areal diversity. It is an interesting empirical question whether the converse of Theorem 1 holds; that is, given an overall Pareto distribution, whether a certain partitioning of the cities will preserve the distribution, after reranking, in each of the partitions. In particular, are cities still R-S distributed within each of the states separately. While this appears likely on the basis of Theorem 1, we have not yet put it to a test.

It is important for most deeper investigations to move from the populations of cities to the underlying variable of population density. For example, it is probably more fruitful theoretically to deal with cities as regions of high population density than with cities as politically incorporated entities with more or less arbitrary boundaries. In particular, this is important with great urban areas, where the cities spill over their political borders into urban fringes, conurbations, satellite cities, etc; and also in rural areas, where the R-S law breaks down and towns begin to merge in the countryside.

There appears to be little known on the important subject of rural population densities. Stewart (Empirical Mathematical Rules...) claims that rural densities are approximately proportional to the square of the quantity he calls the potential of population. The potential at a point is found by summing the reciprocals of the distances of all persons in the country from that point. We have not found any use for this concept so far, and doubt its importance, but we can translate Stewart's rule into our own terms. The New York City region is dominant in the calculation of potential all the way to the Rockies, so that, approximately, the potential of a point is inversely proportional to its distance from New York. This means that the square of the potential is inversely proportional to a circular area surrounding New York with that point on the rim, and so, therefore, is the rural density at that point. But all higher rural densities will tend to fall within this area, and all lower outside it, since these vary positively with potential.

Now think of the country as being chopped up by a gridwork into a great many little regions of equal area. We want to rank these according to density in descending order. We will very soon be rid of all urban densities, since these areas cover something like 2% of the country. When we begin ordering rural densities, we will tend to unwind them from the area surrounding New York. By the time we get to our little area on the rim, we will have counted, approximately, all the little chunks within that circular area around New York. This number will be proportional to that area, since our gridwork was constructed with all chunks of equal area. But we discovered above that, by Stewart's rule, the density of the little rim area was inversely proportional to the circular area.. This means that the density of the little rural chunk is, approximately, inversely proportional to its rank--in other words, that rural densities obey the Rank-Size law! (This might be called the Rank-Density law). So our tortuous (and perhaps torturous) chain of reasoning has carried us from Stewart's rule back to familiar territory.

What about urban densities? Here our theory is in bad shape:

Theorem 2: Assume that cities are R-S distributed, that the density varies as the Qth power of the population of the city, where 1 exceeds Q exceeds zero, and that density is uniform over the area of the city. Then if cities are chopped up by an equi-areal gridwork as above, the chunks will follow the Rank-Size law with small error. Furthermore, the R-S constant for this distribution will be #/Q, where # is the Rank-Size constant for the original city distribution. The proof is omitted as non-edifying.

Stewart claims that the power assumption is good and that the value of Q is about 1/4. Be this as it may, the other assumption, that of uniformity, is definitely bad. Unfortunately, we have not been able to prove the theorem yet under more realistic assumptions...

(Example of a more realistic assumption: that the population density within a city follows a bivariate t-distribution; in particular, a bivariate Cauchy distribution. see below)

If it turns out that population densities follow the Rank-Size rule, as seems probable from the above considerations, the next step would be to turn the trick and derive all other spatial R-S distributions from this one, by the use of Theorem 1 and otherwise.

We turn now to a deeper level of explanation: by migration. The idea is to explain existing distributions as equilibrium patterns of certain migratory propensities.

In this connection we make the following comment on Pareto laws. These are found in a great many queer places; for example, in the distribution of incomes, of words in books, of number of scientific papers published, of firm sizes, of biological species per genus, even perhaps in the distribution of fragment sizes when you smash a beer bottle on the ground. Such diversity is not mysterious, but merely indicates that a fairly simple probability mechanism underlies these manifestations, as Simon has convincingly argued (On a Class of Skew Distribution Functions). For the same reason the normal and Poisson distributions pop up everywhere. But it has become apparent that the probability mechanisms which generate Pareto laws are many and varied. It is likely that different mechanisms are operating in different circumstances. How then are we to find the right one?--by stressing, in the case of city size and other such distributions, exactly what distinguished them from other Pareto distributions, namely, that they are distributed in space as well as size, and that the location of places vis-a-vis each other is of prime importance in establishing their size distribution.

As a preliminary we establish an important condition which is equivalent to the Rank-Size Distribution:

Theorem 3: Divide cities into size strata as follows: take as dividing points populations of sizes x, xy, xy^2,...xy^i,... (x exceeds zero; y exceeds 1). Call the class whose populations lie between xy^{i-1} and xy^i the i-th stratum of cities. Then the fact that the cities follow the Rank-Size law implies that the total populationsof each of the strata are equal to each other. The converse is true if y is close to 1.

Proof: The R-S law states that the number of cities greater than population P is $\#/P$. Then the density of cities at P is $-d(\#/P)/dP$ which equals $\#/P^2$. The number of people in cities with populations between xy^{i-1} and xy^i is, then, $\int_{xy^{i-1}}^{xy^{i}} P(\#/P^2)dP$ which equals $\#\log(xy^i) - \#\log(xy^{i-1})$ which equals $\#\log y$. But this is independent of i, so all strata are equal. QED The converse is omitted.

Theorem 4: Write P_i for the population of the i-th stratum; write prob(m:n) for the probability that a person living in an m-th stratum city will move to an n-th stratum city in unit time. Stratify with y close to 1 (i.e. take very fine strata). Assume an upper and lower bound to city sizes. Then the relation prob(m:n) equals prob(n:m) will generate a Rank-Size distribution.

Indication of proof: The number of people migrating from stratum m to stratum n in unit time is expected to be $P_m$prob(m:n). If P_m exceeds P_n then $P_m$prob(m:n) exceeds $P_n$prob(n:m), since the probabilities cancel; that is, there is always a net expected migration from a greater to a lesser population. Therefore, the most highly populated stratum will be losing population and the least populated will be gaining. The equilibrium obtains only when all populations have the same value, and this is clearly stable. But by Theorem 3 this is equivalent to a Rank-Size distribution. QED (The proof must be modified if some probabilities are zero; the theorem is then false)

This Theorem provides a fundamental probability mechanism for generating R-S distributions, and one that is extraordinarily simple. Much can be done with it.

A Rank-Size distribution in equilibrium conversely implies that prob(m:n) equals prob(n:m), if only we assume there is no "multilateral" equilibrium. This is an eminently reasonable assumption for ordered distributions such as we are dealing with.

We may work Theorem 4 backwards and ask what deviations from the probability rule are implied by observed deviations from the R-S rule. But we have done nothing along these lines as yet.

The preceding theorems go thru identically if we substitute population densities for population.

We come now to the specific form of the migration function. The variables which appear most pertinent in determining the probability that a person will migrate from one location to another in unit time seem to be the population density of destination, the population density of location, and the distance; the probability being positively related to the first and negatively to the last two. (cf. Nelson, Migration, Real Income and Information)

First some comments on the meanings of these terms. Distance should not be measured as the crow flies, but rather functionally, as a measure of the facility of communication between two points. For example, in the old days the distance from New York to California should perhaps have been measured around Cape Horn. Mountains and deserts might give functionally infinite distances at low levels of technology, these shrinking to geographic proportions with airplane. Also, political barriers must be taken as increasing distances. Other examples will occur to the reader. It becomes questionable whether these varying effects can be summed up in a single distance variable. One can only try as a first approximation; in fact, paucity

of data force us to use geographical distances in spite of all.
As to density, we have first of all the familiar distinction between the residential and the occupational city. This can be carried further: we have the recreational city, the week-day vs. week-end city, the day vs. night city, the summer vs. winter city --and so on, for all the rhythms of our round of care. We can grasp these diversities into one concept by defining the density of a location as an appropriate time average of the momentary density. In particular, a location may have a high density even if its population is completely transient. Example: a busy highway in the open country is a location of high density. In fact, such locations attract migrants as any stable population center would--in the form of filling stations, diners and motels. A crossroads location is doubly dense, and is evidently a rather attractive spot. Breaks in transport routes are extremely attractive, a great density resulting from the dead time involved in moving freight from one carrier to another. In fact, it is exactly at such breaks in transport that great cities tend to arise, as Cooley pointed out long ago. The measurement of densities in periodic variation is a deep problem, and in fact a theory is waiting to be created at this very spot.

A migration law based on the three variables is capable of explaining in a rough qualitative way just about all the distributional phenomena we come across which involves only aggregate populations. (That is, excluding differential distributions of different segments of the population, such as segregation, nunneries, old ladies homes, Palm Beach, etc.) For example, it can explain the fanning out of cities along transport routes, the longitudinal distortion of nearby population centers toward each other, the growth of satellite cities, urban sprawl. The crucial test is, however, the quantitative one: can the theory predict the actual distribution of densities in space, and their rates of flow over time?

Let D_{ij} be the distance between points L_i and L_j, and p_i and p_j the respective population densities (these are not identical with P_i and P_j of Theorem 4, first because they are densities, but more importantly, because they refer to individual locations, while P_i and P_j refer to strata distributed over the whole country).

Consider first of all the hypothesis that prob(i:j) is equal to a function of the form $p_i^A p_j^B f(i,j)$, where f(i,j) is a symmetric function of the variables (that is, it equals f(j,i) identically; for example, a constant, or a function involving distance) and A and B are constant. This itself is a strong hypothesis, and furthermore, a readily testable one. We put this in theorem form:

Theorem 5: Let M_{ij} represent the gross migration from location L_i to location L_j. If the law of migration is of the form prob(i:j) equals $p_i^A p_j^B f(i;j)$, where f is symmetric, then for any three locations L_i, L_j, L_k we have $M_{ij}M_{jk}M_{ki}$ equals $M_{ik}M_{kj}M_{ji}$. That is, the product of gross migrationsin the clockwise order equals the product of gross migrations in the counterclockwise order.

Proof: Consider the product prob(i:j)prob(j:k)prob(k:i). It equals $p_i^{A \text{ plus } B} p_j^{A \text{ plus } B} p_k^{A \text{ plus } B} f(i,j)f(j,k)f(k,i)$. But this obviously equals also prob(i:k)prob(k:j)prob(j:i). Multiply both sides of the probability product equality by $p_i p_j p_k$, pair them off with their respective probabilities by origin (e.g. $p_i$prob(i:j) which equals M_{ij}) and the conclusion follows QED

The migration product rule has not yet been tested.

Supposing that the hypothesis has survived this severe test, we pass onward to a more restricted form of the hypothesis. We suppose that prob(i:j) is proportional to $p_i^A p_j^B D_{ij}^C$. Is there anything we can say about the relations among the exponents A, B, and C? There is, given certain very plausible assumptions.

We need, first of all, the assumption of what may be called a Christaller distribution in space of cities (Christaller, Die Zentralen Orte in Süddeutschland as reported in Ullman, A Theory of Location for Cities). This is a roughly regular distribution, with a tendency for cities of like size to distribute themselves in a hexagonal lattice pattern. There need not be the same density of cities thruout the region, but if not the density of all city sizes must go up and down in the same proportion in the various subregions. Further, we require that cities of different sizes be distributed at random relatively to each other. This assumption seems to violate the logic of the Christaller distribution. As a matter of fact, our theorem can follow from precisely the opposite sort of assumption, namely, that all cities are Poisson distributed over the region. This assumption too may be modified by allowing a variable Poisson parameter in different subregions, providing it varies in proportion for all city sizes. The Poisson assumption may turn out to be a good one overall, for while migration tends to make similar densities agglomerate, the Christaller data reveal a tendency for cities to deglomerate.

Next, stratify cities as in Theorem 3 (p.10 above). Choose y close to 1. Assume prob(m:n) equals prob(n:m), which by Theorem 3 is equivalent to Assuming the Rank-Size law. Then

Theorem 6: If the above assumptions respecting the form of the migration law, the modified distribution of cities in space by a Poisson law, and the paired equality of migration probabilities between strata are satisfied, then 2(A-B) equals C.

Indication of proof: A serious notational confusion may arise at this point, because we are dealing with migration probabilities between individual cities and between strata at the same time. We use prob(i:j) to refer to individual city intermigration, and prob(m:n) to refer to stratum intermigration.

The migration from the m-th to the n-th stratum is the summation over all possible pairs of cities, the first member located in the m-th stratum and the second in the n-th. Since y has been chosen close to 1, we may, with small error, set all cities in the m-th stratum at the same population, namely xy^m, and similarly all cities in the n-th stratum at xy^n. Next, distances: we want the mean distance from a given m-stratum city to each of the cities in the n-th stratum. The following theorem may be obtained from probability theory: when points are Poisson distributed in a plane, the mean distance to the k-th nearest point is a certain function of k over the square root of the density. The total population of the n-th stratum is #logy, by the proof to Theorem 3. The number of cities in this stratum is, therefore, $\#\log y/xy^n$. If the total area of the region is E, the density of these cities is $\#\log y/xy^nE$. The mean distance to the k-th nearest point is proportional to $(xy^nE/\#\log y)^{\frac{1}{2}}$. Substituting all this in the migration law we get prob(i:j) equals $(xy^m)^A(xy^n)^B(xy^nE/\#\log y)^{\frac{1}{2}C}$ times some constant depending on k. Add this up over all k; by symmetry, this summation equals prob(m:n). Therefore it equals prob (n:m). But this, conversely, equals a similar summation times over a factor $(xy^n)^A(xy^m)^B(xy^mE/\#\log y)^{\frac{1}{2}C}$ times some constant depending on k. These same constants recur on both sides identically; variations of density will again introduce the same factors on both sides. It follows that the two factors must be equal. Cancelling, this reduces to n(B plus $\frac{1}{2}$C - A) equals m(B plus $\frac{1}{2}$C - A), whence 2(A-B) equals C. QED

Can we infer anything further about the exponents, other than the relation 2(A-B) equals C? There is a fair amount of evidence that C is very close to -2: (a) Nelson's data, while equivocal, indicate that A-B is in the vicinity of -1, from which C equals -2 would follow by Theorem 6; (b) A plausible a priori assignment of values to A and B would be A equals zero, B equals 1; (c) Reilly's law of Retail Gravitation (<u>Methods for the Study of Retail Relationships</u>) --which we have not investigated, states that the breaking point between two shopping centers, where they divide customers equally, is the point where distances squared are respectively proportional to populations of the shopping centers; this implies A equals zero and C equals -2B; (d) telephone toll messages seem to fall off as the square of distance; (e) car and truck trips fall off as the square of distance; (f) marriage rates in the same city fall off about inversely with the distance between residences of spouses, which means they fall off as the square of distance to specific locations (Bossard); (g) Zipf persisted in fitting a P/D element to transport ahd communications data instead of a P/D^2; his graphs invariably had a wider scattering than usual, and the bias in slope away from the desired -1 were in most cases in the direction which would have benefited from a P/D^2 application instead.

On the other side, we know of no data which would seriously embarrass the assumption of an inverse square law for distance migration. Stouffer's data on family movements are equivocal; data for distance of occurrence of news items from place of printing seem to slope the wrong way for a P/D^2 correction(but this is very possibly a case like (f) above, where one expects and inverse D instead; similarly railway express data.

There seems to be some evidence that population densities fall off as $1/D^2$ from a population center. This may be implied by the Rank-Size distribution of population densities (cf. the discussion of Stewart's rule, pp.7-8 above).

The trouble with a $1/D^2$ rule is the singularity which occurs at the origin, which clearly corresponds to nothing real. One possible remedy, the simplest, is to replace the rule by a Cauchy rule: migration propensities, and urban densities go as $1/(a^2$ plus $D^2)$, which leaves things unaffected at a distance while correcting the singularity in a plausible way. The Cauchy distribution has some properties which make it more than an expedient; but this is all speculation.

What remains to be done? Almost everything. We need to look at a wider variety of data, to check our various hypotheses at their testable points. A theory of change over time is almost wholly lacking. The migration concepts have to be amalgamated with vital statistics. International comparisons would prove useful. The problem of compound migrations has to be faced. Correlation with other spatial variables, especially with land rent, income and capital density should prove illuminating. A theory of periodic fluctuations and migration waves might be worked out. All of this hardly scratches the surface when we come to differential distributions (p.12 above). The field is large, and will yield only if laws both close-fitting <u>and</u> simple continue to be discovered. We don't know if this can happen, but at least they should be sought after