New Methods in Statistical Economics

Author(s): Benoit Mandelbrot

Source: *Journal of Political Economy*, Oct., 1963, Vol. 71, No. 5 (Oct., 1963), pp. 421-440

Published by: The University of Chicago Press

Stable URL: http://www.jstor.com/stable/1829014

THE JOURNAL OF
POLITICAL ECONOMY

Volume LXXI OCTOBER 1963 Number 5

NEW METHODS IN STATISTICAL ECONOMICS[1]

BENOIT MANDELBROT
Harvard University and I.B.M. Corporation

This is an informal and non-mathematical presentation of several new mathematical approaches to the study of speculative markets and of other economic phenomena. My principal thesis is that random variables with an infinite population variance are indispensable for a workable description of price changes, the distribution of income, and firm sizes, etc.

A side effect of my work should be a revival of interest in the statistical law of Pareto, which—together with several of its kin—very much dominates the field of economics, and several areas of the physical sciences as well.[2]

I. INTRODUCTION

Neglect commonly marks the attitude of statisticians and of mathematical economists towards Pareto's well-known empirical discovery, that there exist two constants C and $\alpha > 0$, such that the relative number of incomes exceeding u can—for large values of u—be written in the form $Cu^{-\alpha}$. This relation is usually verified by plotting the logarithm of the number of incomes greater than u, as a function of the logarithm of u: one should obtain a curve that for large u becomes a straight line sloping down to the right with an absolute slope equal to α.

It is not very seriously questioned, however, that the law of Pareto represents very satisfactorily, not only the "tail" of the distribution of personal income, but also the tails of the distributions of firm sizes and of city sizes. In fact, the game of searching for new instances of that law has been at times very popular and quite successful, as for example in the hands of George Kingsley Zipf.[3] It therefore seems that the law of

[1] This work was partly supported by the Office of Naval Research, under contract number Nonr-3775(00), NR-047040. I am indebted to many friends—and particularly to Merton Miller and Lester Telser—for detailed discussions of earlier oral and written versions of this paper.

[2] Two privately circulated drafts of this paper were therefore entitled "Towards a Revival of the Statistical Law of Pareto." Sections I–V are a revision of IBM Research Note NC-96: "Aggregation, Choice, Mixture and the Law of Pareto" (May 23, 1962). Section VI is based upon a Harvard talk on "Illusory Regularities, History and Forecasting" (December 6, 1962).

[3] G. K. Zipf, *National Unity and Disunity* (Bloomington, Ind.: Principia Press, 1941), and his *Human*

Pareto has been neglected because it does not represent the middle range of incomes—which may be the more important for certain purposes—and also because it lacks theoretical justification—at least within the context of elementary probability theory. I believe, however, that this remarkable finding deserves a systematic new examination in the light of new methods that I have attempted to introduce into statistical economics. I shall indeed show that *the law of Pareto literally thrusts itself* upon anyone who takes seriously the simplified models based on maximization or upon linear aggregation, upon anyone who takes a cautious view of the origin of the economic data, and upon anyone who believes in the influence on economics of the physical distribution of various scarce mineral resources and of rainfall.

I shall also show the following: when the "spontaneous activity" of a system is ruled by a Paretian process, the causally structural features of the system are likely to be very much more hidden by noise than is the case where the latter is Gaussian. Causal structures may even be totally "drowned out." On the other hand, Paretian noise generates all kinds of "patterns" that seem to be perfectly clear-cut but have no value for purposes of prediction. Thus, in the presence of a Paretian "spontaneous activity," the scientist is faced by an unexpectedly heavy burden of proof, and the basic problem of validation of laws acquires many new and quite perturbing features.

We shall see that the most important features of the law of Pareto are linked to the length of its tail, and not to its extreme skewness. In fact, in the case of random variables that can also take large negative values, one must also introduce various bilaterally Paretian distributions, which may even be symmetric. Hence, the extreme skewness of the distribution of income must be considered as being a secondary feature of those Paretian variables that are constrained to be positive.

To terminate this introduction, I shall state according to which criteria I hope that one will judge the research presented in this paper: this work is not primarily mathematical, since many of the tools it uses have long been available. Nor is it primarily an account of empirical findings, even though I was presumably the first to establish some important properties of temporal changes of speculative prices. The best is perhaps to say that I hope that several methods that I have proposed will prove to constitute workable "keys" to further developments along a long-mired frontier of economics. Their value should rest: (1) on the length and number of the successful chains of reasoning which they have made possible; (2) on the number of seemingly reasonable questions that they may show to be actually "ill-set" and hence without answer; (3) last but of course not least, on the practical importance of the areas in which all these developments take place.

I shall not attempt in this paper to treat any point exhaustively, nor to specify fully all the conditions of validity of my assertions, which are discussed in detail in the publications referred to in various footnotes. Many readers may prefer to read Section VI before Sections II–IV. Section IX examines an important and influential critique of the law of Pareto.

Behavior and the Principle of Least-Effort (Cambridge, Mass.: Addison-Wesley Press, 1949).

II. THE GENERAL PRINCIPAL OF MY "METHOD OF INVARIANT LAWS"

The approach used in my studies of the law of Pareto may seem unusual in the context of social science and it arose from a method familiar in physics[4] (but this fact is of no concern to the economist). To begin with, I believe that the various "microscopic models," which could be considered as explaining "why" such and such a version of the law of Pareto is encountered in such and such a domain, are ordinarily less convincing than the law itself, because they are of less general applicability, and because seemingly slight and irrelevant changes in the hypotheses often completely change the result.[5] But the economist need not agree with my evaluation of the existing generative models, or he may hope that future ones will be better. Whatever his opinion of my motivation, it is also clearly important to examine carefully under which conditions empirical observation is actually practiced in economics and in other social sciences. By "observation" one should not only mean the activity of the scholar who observes to describe, but also that of the entrepreneur who observes to act. In both cases, it is clear that most economic quantities can hardly ever be observed directly, and are usually altered by some ill-known sequence of all kinds of manipulation.

In most practical problems, very little can be done about this difficulty, and one must make do with whatever approximation to the desired data is actually available. But inappropriate data are a notorious handicap in theoretical work, since economic relationships usually relate to conceptually irreducible economic quantities and cannot generally be expected to be left invariant by the manipulations performed before actual measurement. That is, the analytical formulas by which economic relationships may be described must be expected to change markedly in form whenever one applies one of the basic transformations. As a result, however great the practical importance of these relationships, and hence however great the efforts to understand them, there is a good chance that their form will be discovered later, and that they will forever remain known with lesser precision, than the phenomena that "happen" to be in some sense invariant with respect to the maximum number of observational transformations such as the following (which are all fundamental, but unequally so).

Linear aggregation, or simple addition of various quantities in their common natural scale.—For example, aggregates of various kinds of income are better known than each kind taken separately. Long-term changes in most economic quantities are better known than the more desirable medium-term changes; moreover, the meaning of "medium term" differs between series, so that a law that is not invariant under aggregation would be apparent in some series and not in others, and could not be firmly established. A number of operations of aggregation also occur in the context of firm sizes, in particular when "old" firms merge within a "new" one.

The most universal type of aggregation occurs, however, in linear models

[4] However, the various criteria of invariance used by physicists are somewhat different in principle. For example, the classical "principle of relativity" was not introduced to explain any complicated empirical law such as that of Pareto. I am indebted to Harrison White for the suggestion to stress the nuances between my methods and those of physics.

[5] This is stressed in my "Survey of the Growth and Diffusion Models of the Law of Pareto," a privately circulated "Research Note," to appear.

that add the (weighted) contributions of several "causes," or more generally embody linear relationships between several variables or between the current and the past values of a single variable (autoregressive schemes). Everybody's preference for such models is of course based upon the unhappy but unquestionable fact that mathematics offers few workable non-linear tools to the scientist.

There is actually nothing new in my emphasis upon invariance under aggregation. It is indeed well known that the sum of two independent Gaussian variables is itself Gaussian, and this—after ease of analytical manipulation—is the principal reason for using Gaussian "error terms" in linear models. However, the common belief that only the Gaussian is invariant under aggregation is correct *only* if one excludes random variables with infinite population moments, which I shall *not* do (see Sec. V). (Besides, the Gaussian law is *not* invariant under our other two observational transformations.)

Note also that one may aggregate a small or a very large number of quantities. Whenever possible, "very large" is approximated by "infinite," so that aggregation is intimately related with the question of the central limit theorems concerning the behavior of limits of weighted sums of random variables.

Weighted mixture.—For example, a weighted lottery ticket would be one in which a first preliminary chance drawing would determine in which of several final drawings the gambler has the right to participate. This provides a model for all kinds of actually observed variables. For example, if one does not know the precise origin of a given set of income data, one may consider that it was picked at random among a number of possible basic distributions; the distribution of observed incomes would then be a mixture of the basic distributions. Similarly, price data often refer to grades of a commodity that are not precisely known and hence can be assumed to be randomly determined. Finally, the very notion of a firm is to some extent indeterminate, as one can see in the case of almost wholly owned, but legally distinct, subsidiaries, and available data often refer to "firms" that actually vary in size between individual establishments and holding companies; such a mixture may be represented by random weighting.

In many cases, one deals with a combination of the above operations: for example, after a wave of mergers has hit an industry, one may consider that the distribution of "new" firms is a *mixture* of the distribution of companies *not* involved in a merger, and of the distribution of companies made up of the *sum* of two old firms, and perhaps even of *sums* of more than two firms.

Maximizing choice, the selection of the largest or smallest quantity in a set. For example, it may be that all we know about a set of quantities is the size of the one chosen by a profit maximizer. Similarly, if one uses historical data, one must often expect to find that the only fully reported events are the exceptional ones, such as droughts or floods, famines (and the names of the "bad kings" who reigned in those times), or "good times" (and the names of the "good kings"). Worse still, many data are a mixture of full reporting and of reporting limited to the extreme cases.

Although the above transformations are not the only ones of interest, they are so important that one should characterize the laws that they leave invariant It so happens that *invariance-up-to-scale holds asymptotically for all three transformations if the parts follow the law of Pareto* (in the case of infinite aggregation, invariance only holds if Pareto's exponent is less than two). On the contrary (with some qualifications) *invariance does not hold—even asymptotically—in any other case.* Hence, if one's belief in the importance of those transformations has any strength at all, one will attach a special importance to Paretian phenomena, at least from a purely pragmatic viewpoint.

This proposition also affects the proper presentation of empirical results. Indeed, one knows that in order to be precise in the statement of scientific laws, it is *not* sufficient to say that income, for example, is Paretian; it is also necessary to list the excluded alternatives. My considerations will show that the proper precise statement *is not of the form:* "It is true that incomes (or firm sizes) follow the law of Pareto; it is not true that incomes follow either the Gaussian, or the Poisson, or the negative binomial or the log-normal law." *We must rather say:* "It is true that incomes (or firm sizes) follow the law of Pareto; it is not true that the distributions of income are very sensitive to the methods of reporting and of observation."

III. SOME INVARIANCE PROPERTIES OF PARETO'S LAW AND OF CERTAIN OF ITS KIN

Of course, the singular character of the asymptotic law of Pareto holds only under additional assumptions, so that the problem will surely not be exhausted by the present approach. Consider indeed N independent random variables, $U_n (1 \leq n \leq N)$, following the weak (asymptotic) form of the law of Pareto, with *the same exponent* α, so that

$$Pr(U_n > u) \sim C_n u^{-\alpha} \text{ if } u \text{ is large .}$$

The behavior of $Pr(U_n < -u)$ for large u will be examined in Section VII.

Keeping the proofs in footnotes, let me begin by quoting some statements that imply that a Paretian behavior of U_n is *sufficient* for the three types of asymptotic invariance—up-to-scale. The sign Σ will always refer to the addition of the terms relative to the N possible values of the index n.

Weighted Mixture.—Suppose that the random variable U_W is a weighted mixture of the U_n, and that it has the probability p_n of being identical to U_n. One can show that this U_W is also asymptotically Paretian, and that its scale parameter is $C_W = \Sigma p_n C_n$, which is simply the *weighted average* of the separate scale coefficients C_n.[6]

Maximizing choice.—Let U_M be the largest of the variables U_n, that is, the one that turns out a posteriori to be the largest, when the values of all the U_n are known. One can show that this U_M is also asymptotically Paretian, with the scale parameter $C_M = \Sigma C_n$ which is the *sum* of the separate scale coefficients C_n.[7]

Aggregation.—Let U_A be the sum of the random variables U_n. One can show that it is also asymptotically Paretian, with a scale parameter that is again the *sum* of the separate C_n.[8] Thus, the sum

[6] *Mixture.*—It is easy to see that one has

$$Pr(U_W > u) = \Sigma p_n Pr(U_n > u) \sim \Sigma C_n p_n u_n^{-\alpha} = C_W u^{-\alpha} .$$

[7] *Maximizing choice.*—In order that $U_M \leq u$, it is clearly both necessary and sufficient that $U_n \leq u$ *for every n.* Hence, using Π to designate the product of the terms relative to the N possible values of the index n, we have

$$Pr(U_M \leq u) = \Pi Pr(U_n \leq u).$$

It follows that one has:

$$Pr(U_M > u) = 1 - Pr(U_M \leq u) \sim 1 - \Pi(1 - C_n u^{-\alpha}) \sim \Sigma C_n u^{-\alpha} = C_M u^{-\alpha} .$$

[8] *Aggregation.*—Here the argument is more involved, and I prefer to suggest that the reader look

of the U_n behaves asymptotically exactly like the largest of them.

Mixture combined with aggregation is an operation that occurs in the theory of random mergers of industrial firms.[9] One can show that it also leaves the law of Pareto invariant up to scale.

The converses of the above statements are true only in the first approximation: in order for the invariances-up-to-scale to hold, the distributions of the U_n need not strictly follow the law of Pareto; but they must be so close to it as to be practically Paretian.

Strictly invariant and limit distributions.—Let me now abandon asymptotics and introduce Fréchet's and Lévy's kins of the law of Pareto, by imitating (with a different interpretation) a famous principle of physics: to require that the random variables U_n be strictly invariant up to scale with respect to one of our three transformations. This means the following: let the N random variables U_n all follow—up to changes of scale—the same law as the variable U, so that U_n can be written as a_nU, where $a_n > 0$. I shall require that U_W (respectively U_M or U_A) also follow—up to changes of scale—the same law as U, which means that one can write U_W (respectively U_M or U_A) in the form a_WU (respectively a_MU or a_AU), where a_W, a_M and a_A are some positive functions of the numbers a_n.

It turns out that the conditions of invariance lead to somewhat similar equations in all three cases;[10] ultimately, one obtains the following results:

Maximization.—The invariant laws must be of the form $F_M(u) = \exp(-u^{-\alpha})$; this result is due to Maurice Fréchet.[11] They are clearly Paretian, since—for large u—F_M can be approximated by $1 - Cu^{-\alpha}$. They also "happen" to have the remarkable property of being the limit distributions of expressions of the

---

up the proof, for example in my "The Pareto-Lévy Law and the Distribution of Income," *International Economic Review*, I (May, 1960), 79–106.

[9] See my "Oligopoly, Merger and the Paretian Size Distribution of Firms," a privately circulated "Research Note" to appear.

[10] Let U be characterized by its distribution function $F(u) = Pr(U \leq u)$ and by its generating function $G(s)$, which is the Laplace transform of $F(u)$: $G(s) = \int_{-\infty}^{\infty} \exp(-us)dF(u)$. (This limits our argument to laws for which dF is so small for $u \to -\infty$ that G converges.) Then, one can begin by writing the following conditions, which are respectively necessary for the various types of invariance—up-to-scale.

Weighted Mixture.—It is necessary that stability hold for equal p_n. Thus, it is in particular necessary that the function F satisfy the condition that

$$\frac{1}{N} \Sigma F(u/a_n) = F(u/a_W).$$

Maximization.—Now, it is necessary that $F(u/a_M) = \Pi F(u/a_n)$; in other words, one must have:

$$\Sigma \log F(u/a_n) = \log F(u/a_M).$$

Aggregation.—It is here necessary that

$$\Sigma \log G(a_n s) = \log G(a_A s).$$

It turns out therefore that the three types of invariance lead to formerly almost identical equations, although they refer to different functions, respectively F_W, $\log F_M$ and $\log G_A(s)$. The general solutions must therefore respectively take the forms $F_W(u) = C' - Cu^{-\alpha}$; $F_M(u) = \exp(-Cu^{-\alpha})$ and $G_A(s) = \exp(-Cs^{-\alpha})$. One also easily verifies that $a_M^{\alpha} = a_A^{\alpha} = \Sigma a_n^{\alpha}$; $a_W^{\alpha} = (1/N)\Sigma a_n^{\alpha}$.

I shall now show that the above necessary conditions are actually not sufficient, and that additional requirements must be imposed upon C', C and α.

Maximization.—The distribution function of a random variable must be non-decreasing and such that $F_M(\infty) = 1$. This requires that $C > 0$ and $\alpha > 0$, which leaves us with the laws $F_M(u) = \exp(-Cu^{-\alpha})$.

Mixture.—In order that $F_W(u)$ be non-decreasing and such that $F_W(\infty) = 1$, it is now necessary that $C' = 1$, $\alpha > 0$ and $C > 0$.

Aggregation.—In order that $G_A(s)$ be a generating function, one can show that it is necessary that $0 < \alpha < 1$ with $C < 0$, or $1 < \alpha \leq 2$ with $C > 0$.

[11] M. Fréchet, "Sur la loi de probabilité de l'écart maximum," *Annales de la Société Polonaise de Mathématiques* (Cracovie), VI (1927), 93 ff.; see also E. Gumbel, *Statistics of Extremes* (New York: Columbia University Press, 1958).

form $N^{-1/\alpha} \max U_n$, where the U_n are asymptotically Paretian. There are no other distributions that can be obtained simply by multiplying $\max U_n$ by an appropriate factor and by having N tend to infinity. (However, if one also allows the origin of U to change as $N \to \infty$, one can obtain the "Fisher-Tippett distribution," which is *not* Paretian and is not invariant under the other two transformations.)

Mixture.—In this case, invariance leads to $F_W(u) = 1 - Cu^{-\alpha}$, that is, to the analytical form of the law of Pareto extended down to $u = 0$. Such a solution corresponds to an infinite total probability, so that, strictly speaking, it is inacceptable. However, it must not be rejected out of hand, because in many cases in practice U is further restricted by some relation of the form $0 < a \leq u \leq b$, leading to a perfectly acceptable conditional probability distribution.

Aggregation.—Finally, aggregation leads to random variables that are the "positive" members of the family of Lévy's "stable distributions," other members of which will be encountered later.[12] They depend upon several parameters, the principal of which is again designated by α and must be such that $0 < \alpha \leq 2$.

One knows $dF_A(u)$ in closed analytic form for the stable law with $\alpha = 2$, which is the Gaussian (in other words, this classical law is in a sense a limit case of the stable laws, but is not itself Paretian). One also knows the positive stable law with $\alpha = \frac{1}{2}$, which plays a central role in return to equilibrium in coin tossing. In other cases, no closed analytic expression is known for the stable $F_A(u)$; Lévy has shown, however, that they asymptotically follow the law of Pareto of exponent α, except if $\alpha = 2$ (for α just below 2, they tend slowly to their Paretian limit).

The stable variables yielded by the present argument can take negative values if $1 \leq \alpha \leq 2$, as is readily seen in the Gaussian case. But there is a very small probability that they take *large* negative values, and I have shown how this question can be handled in practice with the help of appropriate changes of origin.

Lévy's stable distributions have another important property: they are the only possible non-Gaussian limits of linearly weighted sums of random variables. Hence, even though they cannot begin to compare with the Gaussian law from the viewpoint of ease of mathematical manipulation, they share both the fundamental properties of that law from the viewpoint of linear operations. The existence of the corresponding forms of the non-classical central limit theorem show that, if a process is the resultant of many additive contributions, it need *not* be Gaussian; if one wishes to explain by linear addition a phenomenon that is ruled by a skew distribution, it is *not* necessary to assume that the addition in question is performed in the scale of log U rather than in the scale of U itself. This also shows that the log-normal distribution is *not* the only skew law that can be explained by addition arguments, thus taking away the principal asset of that law (which is known in most cases to underestimate grossly the largest values that can be taken by the variable of interest).

One can see that the probability densities of the three invariant families differ through most of the range of u. However,

[12] Paul Lévy, *Calcul des probabilités* (Paris: Gauthier-Villars, 1925), and B. V. Gnedenko and A. N. Kolmogoroff, *Limit Distributions for Sums of Independent Random Variables*, trans. K. L. Chung (Reading, Mass.: Addison-Wesley Press, 1954).

if $0 < \alpha < 2$, their asymptotic behaviors coincide, so that the law of Pareto is also asymptotically invariant with respect to applications of an arbitrary succession of the basic transformations (if α is close to 2, the practical application of this property requires additional qualifying statements).

It should be noted that Fréchet's and Lévy's Paretian limit distributions have attracted substantial attention from pure mathematicians. However, the generally known applications of Paretian maximum distributions were few and those of Paretian sum distributions (stable laws) were practically non-existent. It is true that the introduction to the celebrated Gnedenko-Kolmogoroff treatise[13] contains claims concerning the wide applicability of the mathematical techniques to which that book is devoted and even references to forthcoming publications specially concerned with applications. However, when I discussed this introduction with Kolmogoroff in 1958 (ten years after the appearance of the original Russian edition), I found that these papers had not materialized after all—for lack of applications! Basically, the only fairly well-known practical instance of a stable distribution remains the law, due to Holtsmark but often rediscovered, which rules the Newtonian attraction between randomly distributed stars.[14] Thus, the Gnedenko-Kolmogoroff book has not pre-empted my plea that stable laws be counted among the most "common" probability distributions.

IV. ON THE VALUE OF THE EVIDENCE OF DOUBLY LOGARITHMIC GRAPHS

Limitations on the value of α lead to another quite different aspect of the general problem of observation relative to the practical significance of statements having only an asymptotic validity. Indeed, in order to verify empirically the law of Pareto, the usual first step is to draw the so-called doubly logarithmic graph of $\log_{10} [1 - F(u)]$ as a function of $\log_{10} u$. One should find that this graph is a straight line with the slope $-\alpha$, or at least that it rapidly becomes straight as u increases. But let us look closer at the sampling point of largest u. Except for the distributions of incomes, one seldom has samples of over 1,000 or 2,000 items; or one may otherwise know the value of u that is exceeded with the frequency $1 - F(u) = 1{,}000^{-1}$ or $2{,}000^{-1}$. That is, the "height" of the sampling doubly logarithmic graph will seldom exceed *three* units of the decimal logarithm of $1 - F$. The "width" of this graph will therefore be at the very best equal to $3/\alpha$ units of the decimal logarithm of u. However, if one wants to estimate reliably the value of the slope α, it is necessary that the width of the graph be close to one unit: therefore, one cannot have any trust whatsoever in data that suggest that α is larger than 3, and the practical range of α's is anyway hardly wider than in the case of Lévy's Paretian laws.

Looking at the same question from another angle, consider a Gaussian, log-normal, negative binomial or exponential distribution on doubly logarithmic paper: since these distributions are all very "short-tailed," the slope of the graph will become asymptotically infinite. However, in the region of probabilities down to $1{,}000^{-1}$, the dispersion of sample data is likely to generate—on doubly logarithmic co-ordinates—the appearance of a straight line having a high but finite slope. In the words of F. Macaulay (see Sec. IX): "The approximate linearity of

[13] Gnedenko and Kolmogoroff, *op. cit.*, n. 12.

[14] See Section 2.8 of my "The Pareto-Lévy Law and the Distribution of Income," *op. cit.*, n. 8.

the tail of a frequency distribution charted on a double logarithmic scale signifies relatively little, because it is such a common characteristic of frequency distributions of many and various types." However, linearity with a low slope signifies a great deal indeed. Figure 1 further illustrates this difference between different values of α.

There is another way of describing curve-fitting using special papers: one may say that the maximum distance between the sample curve and some reference curve—preferably a straight line—defines a kind of "distance" between two alternative probability laws. Any special paper—whether it be log-normal or Paretian—should be used only in ranges where the distances which it defines are sensitive to the differences that count from the viewpoint of the problems at hand. Hence, the conservative thing to do is often to consider several hypotheses, that is, to use several kinds of paper.

To sum up, if one takes account of mixtures, maximization, and practical measurement, the range of values of α is reduced to the interval from 0 to 3. If one also takes account of aggregation, α must fall between 0 and 2 (actually, the range of "apparent" α's is somewhat wider).

V. THE PROBLEM OF THE MEANING OF RANDOM VARIABLES WITH INFINITE POPULATION MOMENTS

Paretian laws are extraordinarily long-tailed, as measured by Gaussian standards. In particular, if $\alpha < 2$, the population second moment is infinite. It should be stressed, however, that there is nothing improper in the concept of such a distribution.

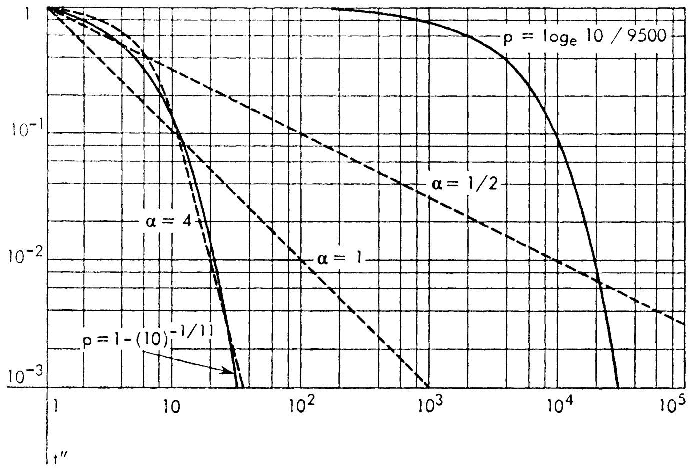

FIG. 1.—Five doubly logarithmic plots: (*A*) Two exponential distributions (*very curved solid lines*), having very different means. (*B*) Two distributions satisfying Pareto's law from $u = 1$ on, and having the exponents $\frac{1}{2}$ and 1. (*C*) Distribution having asymptotically a Paretian exponent of 4. I hope that the relations between these laws demonstrate graphically that distributions similar to (*C*) can readily be confused with the exponential but that small α exponents are reliable.

It is of course true that—observed variables being finite—the sample moments of all orders are themselves finite for finite sample sizes; but this does not exclude that they tend to infinity with the sample size. It may also be true that the asymptotic behavior of samples is practically irrelevant, because the sizes of all empirical samples are by nature finite. For example, one may argue that the history of speculative cotton prices is mostly a set of data from 1816 to 1958, because speculation on cotton was very much diminished by the 1958 acts of the Congress of the United States. Similarly, when one studies the sizes of United States cities, one deals with statistical populations for which the sample size is bounded. Even for continuing series, one may well argue for "après moi, le déluge," and neglect any time horizon longer than a man's life. Hence, the behavior of the moments for infinite sample sizes may seem unimportant. But all that this actually implies is that the only meaningful consequences of infinite population moments are those relative to the sample moments of increasing *subsets* of our various bounded universes.

Figure 2 illustrates what one should expect then from the mathematical viewpoint. I have plotted there the computer simulations ("Monte Carlo experiments") of the variation of the first and second moments corresponding to distinct samples of Paretian random variables with $\alpha = 1$, obtained by simply inverting series of random variables distributed uniformly over the interval (0 1). The second sample moments illustrate what happens when the population mo ment is given by a rapidly divergent integral; the first sample moments illustrate what happens when the population moment is given by a barely divergent

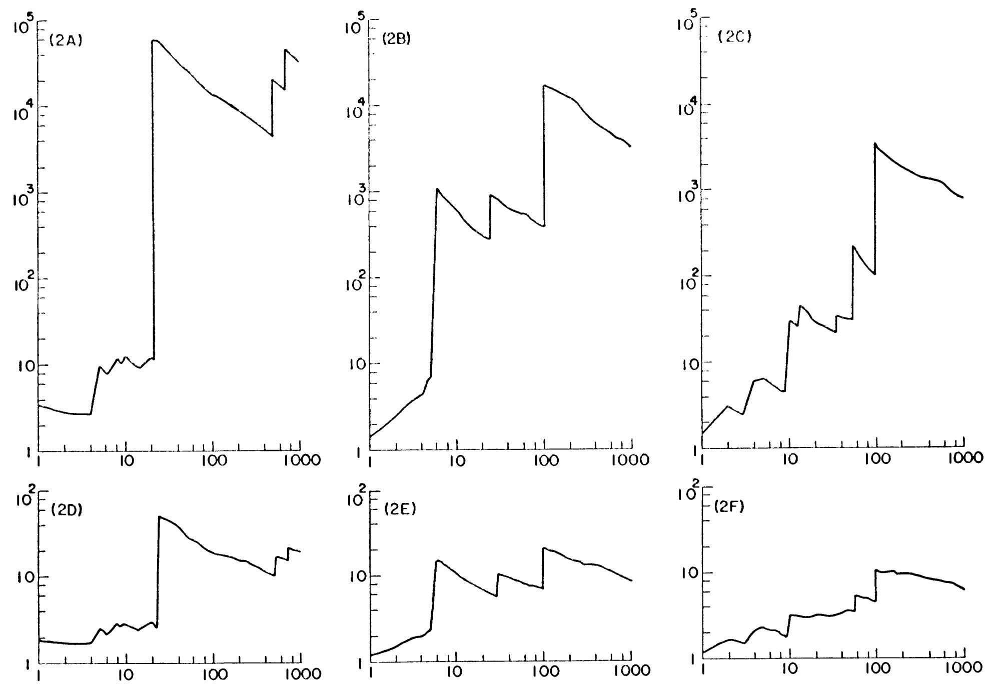

FIG. 2.—Monte Carlo runs of the sequential first moment (*lower graphs*) and the sequential second moment (*upper graphs*) of three independent samples from a Paretian population of exponent 1. The term "sequential moment" means that, in each run, the moment is computed for every sample size from 1 to 1,000. I hope that this figure gives an idea of how erratic and sample-dependent the moments of Paretian variables can be expected to be.

integral. We see that the rapidity of growth of sample moments is even less impressive than their erratic and very sample-dependent character.

Let us now return to experimental data. There is no question that, whenever the sample second moment is observed to "stabilize" rapidly around the value corresponding to the total set, it is useful to take that value as an estimate of the population second moment of a conjectural infinite population from which the sample could have been drawn. But suppose that, as in the case illustrated in Figure 3, the sample second moments corresponding to increasing subsets continue to vary widely even when the sample size approaches the maximum imposed by the subject matter. From the viewpoint of sampling, this must be interpreted as meaning that the distribution is such that even the largest available sample is too small for reliable estimation of the population second moment, or—in other words—that a wide range of values of the population second moment are equally compatible with the data. Suppose moreover—as in Figure 3—that the appearance of the sample data recalls Figure 2. Then it will frequently turn out that the reasonable range of values for the population moment will include the value "infinity," implying that facts can be equally well described by assuming that the "actual" moment is finite but extremely large, or by assuming that it is infinite.

In order to support the alternative that I prefer, let me point out that a realistic scientific model must not depend too critically upon quantities that are difficult to measure. The finite-moment model is unfortunately very sensitive to the value of the population second moment, and there are many other ways in which the first assumption, which of course is the more reasonable a priori, also happens to be by far the more cumbersome analytically. The second assumption on the contrary leads to simple analytical developments, and the

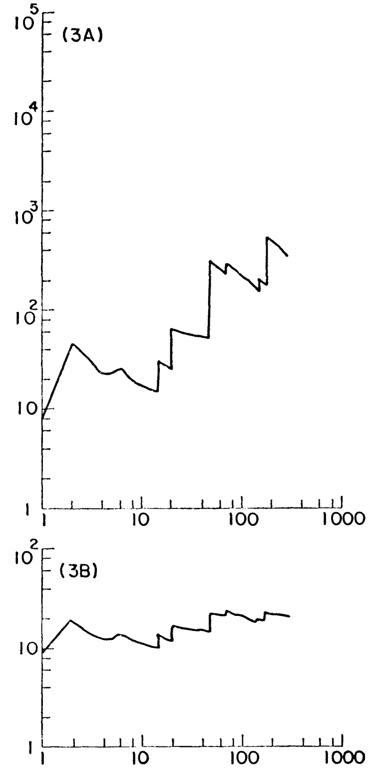

FIG. 3.—United States cities of over 50,000 inhabitants having been ordered alphabetically, this is the record of the sequential first moment (*lower graph*) and of the sequential second moment (*upper graph*) of their populations. These city sizes turn out to have a Paretian exponent of 1.1 or so, so that the first sample moment tends—very slowly—to a limit, while the second sample moment increases less rapidly than on the simulations reported on Figure 2.

rapidity of growth of the sample second moment can be so adjusted that it would lead to absurd results only if one applied it to "infinite" samples, that is, if one raised problems devoid of concrete meaning.

In other words, there is nothing absurd

in assuming, as I am constantly led to do, that intrinsically bounded variables are drawn at random from infinite populations of unbounded variables having an infinite second moment. But all those infinities are a relative matter, entirely dependent upon the statistician's span of interest; as the maximum useful sample size increases, the range of the estimates of the second moment will steadily narrow. Hence, beyond a limit, the second moments of some variables may have to be considered as actually being finite; conversely, there are variables for which the second moment must be considered as being finite only if the useful sample size is smaller than some limit.

Actually, such a use of infinity is a most common one in statistics, insofar as it concerns the function $\max(u_1, u_2, \ldots u_N)$ of the observations. From this viewpoint, even the use of infinite spans would seem to be improper; however, it is well known in statistics that little could be done if one could not use unbounded variables: one even uses the Gaussian to represent the height of adult humans, which is surely positive!

The unusual behavior of the moments of Paretian distributions can be used to introduce the least precise interpretation of the validity of the law of Pareto. For example, if the first moment is finite, but the second moment is infinite, the function $1 - F(u)$ must decrease slower than $1/u^2$ but faster than $1/u$ when u tends to infinity. In this case, the behavior of $F(u)$ in the tails is very important, and, in the first approximation, it may be very useful to approximate it by the form $Cu^{-\alpha}$, with $1 < \alpha < 2$; this can never lead to harm, as long as one limits oneself to consequences that are not very sensitive to the actual value of α. If on the contrary the tail is very short (say if moments are finite up to the fourth order), the behavior of the function $F(u)$ for large u is far less important to represent than its behavior elsewhere; hence, one will risk little harm with interpolations by the Gaussian or the log-normal distribution.

VI. PROBLEMS OF STATISTICAL INFERENCE AND OF CONFIRMATION OF SCIENTIFIC LAWS, WHEN "BACKGROUND NOISE" IS PARETIAN

It is well known that second moments are heavily used in statistical measures of dispersion or of "standard deviation" and in "least-squares" and "spectral" methods. Hence, whenever the considerations of Section V are required to explain the erratic behavior of sample second moments, a substantial portion of the usual methods of statistics should be expected to fail. Examples of such failure have of course often been observed empirically, and have perhaps contributed to the disrepute in which many writers hold the law of Pareto; but it is clearly unfair to blame a formal expression for the complications made inevitable by the data which it represents. If $2 < \alpha < 3$, second moments exist, but concepts based upon third and fourth moments, such as Pearson's measures of skewness and of kurtosis, are meaningless.

I am sure that for practical purposes some of those difficulties will eventually be solved. However, as of today, they are so severe as even to require a re-examination of the meaning of the popular but vague concept of "a structure." It is indeed a truism for the working scientist, especially in fields where actual experimentation is impossible, that the major danger of his trade is the possibility of confusion between patterns that can only be used for "historical" description of his records and those that are also useful for forecasting some aspects of the future. In particular, as we have seen, modern inference theory has taught us always to list both the accepted and the rejected possibilities, and the scientists' major problem is frequently to determine whether a conjectured "relation" is significant with respect to what may be generally called "spontaneous activity," which is the resultant of all the influences that one cannot or does not want to control in the problem at hand, and which is conveniently described with the help of various stochastic models. A useful vocabulary considers the search for laws as a kind of extraction and identification of a "signal" in the presence of "noise."

It is not enough however that all the members of a cultural group agree upon the patterns that they read into a historical record. Indeed, although there is unanimity in the interpretation of *certain* of Rorschach's inkblots, they have no significance from the viewpoint of science as a system of *predictions*. Broadly speaking, a pattern is scientifically significant and is felt to have chances of being repeated, only if in some sense its "likelihood" of having occurred by chance is very small. This kind of significance is obviously to be assessed with the help of the tools of statistics; unfortunately, those have been mostly designed to deal with Gaussian alternatives and, when the chance alternative is Paretian, they are not conservative or "robust" enough *by far*. I believe that one will often be able to get around this difficulty. But, when one works in a field where the background noise is Paretian, one must realize that one faces a burden of proof that is closer to that of history and autobiography than to that of physics.

The same thought can be presented in more optimistic-sounding words, by saying that if a "mere chance" can so readily be confused with a causal structure, it is itself entitled to the same noble designation; the usual word "noise" may perhaps be reserved for the Gaussian error terms, or its binomial or Poisson kins, which are seldom respected as sources of anything that looks interesting.

The situation is made worse by the fact that, in models known to be very structured (for example, to be autoregressive) with a Paretian noise, one should expect the generated paths to be *much* more influenced by the noise, and much less influenced by the structure, than in the Gaussian case—where noise, is already very influential.

The association between the law of Pareto and "interesting patterns" is nowhere more striking than in the outcome of accumulated tosses of a coin. Indeed, the following fact is examined in the later parts of most good books on probability: suppose that we break into the game of tossing a fair coin, which Peter and Paul have been playing since sometime in the early eighteenth century: Whenever the coin falls on "heads," Peter wins a dollar (or perhaps rather a thaler); whenever the coin falls on tails, Paul wins. Let T designate the time it takes for Peter and Paul's fortunes to return to the state they were in at the moment when we broke in. For large values t of T, one has the relation:[15]

$$\left\{\begin{array}{l}\text{Probability that the fortunes return}\\ \text{to their initial state after a time}\\ \text{greater than } t = \text{(constant) } t^{-1/2},\end{array}\right.$$

which is the law of Pareto with an exponent of $\frac{1}{2}$.

However, it is notorious that gamblers see an enormous amount of interesting

[15] The standard treatise on coin-tossing is W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. I, 2d ed. (New York: John Wiley & Sons, 1957).

detail in the past records of *accumulated* coin tossing gains (even more than in the separate results of tossing a coin), and that many are prepared to risk their fortunes on the proposition that these details are not due to mere chance. Several of my papers were originally based upon the idea that very similar phenomena ought to be expected whenever the law of Pareto applies: If so, one could associate with those phenomena some stochastic models that dispense with any kind of built-in causal structure, and yet generate paths in which both the unskilled and the skilled eye distinguish the kind of detail that is usually associated with causal relations. Similar details would be so unlikely in the path generated by a Gaussian process, that they would surely be considered as significant for forecasting; but this is not so in the Paretian case. From the viewpoint of prediction, those structures should be considered as being *perceptual illusions:* they are in the observer's current records and in his brain, but not in the mechanism that has generated these records and that will generate the future events.

Bearing in mind the existence of such models, let us suppose that we have to infer a process from the data. I believe that, in many cases, a non-structured Paretian universe is capable of accounting so well for the observations that it will be extremely difficult at best to choose between alternative models, one of which consciously imbeds some causal relations, while the other has no structure other than stochastic. A student's belief in the existence of "genuine" structures will therefore be challengeable only with the greatest difficulty; conversely, in order to communicate such a belief to others, with the standards of credibility current in physical science, one will need *much* more than the tests of significance that some social scientists shrug off at the end of a discussion. Such a situation will require a drastic sharpening of the distinction between patterns that—however great the scholar's diligence—can only be useful for historical purposes, and those useable for forecasting the future.

The question I have in mind can be well illustrated by the problem of the significance of "cycles." Either by attentively looking at many charts or by using the most sophisticated methods of Fourier analysis, it is comparatively easy to show that almost any record of the past is made up of some combination of swings. But the same is also true for a wide variety of artificial series generated by random processes with no built-in cyclic behavior whatsoever; and it is known that, however great their skill, cycle researchers seldom risk firm short-term forecasts. Could we then ask, paraphrasing Keynes's comments upon early econometric models, "How far are these curves meant to be no more than a piece of historical curve-fitting and description, and how far do they make inductive claims with reference to the future as well as the past?"

It may also be noted that, because of the invariance of the law of Pareto with respect to various transformations (Sec. III), one cannot hope that a simple way out will be provided by arguing that only the genuine structures will be apparent to all observers. The only criterion of trustworthiness is replicability in time.

In an important way, the models of Paretian spontaneous activity diverge from the standards of "operationalism" suggested by philosophers. Indeed, in order to explain by mere chance any given set of phenomena, it will be necessary to imbed them in a universe that also contains such a fantastic number of other possibilities, that billions of years may be necessary to run through all of them. Hence, within our lifetime, any given configuration will occur at most once and one could hardly define a probability for them on the basis of sample frequency. This conceptual difficulty is of course common knowledge among physicists, and it is to be regretted that the philosophical discussions of the foundations of probability so seldom investigate this point. In a way, the physicists' models freely indulge in practices that for the historian are mortal sins: to rewrite history as it would have been if Cleopatra's nose had had a different shape. My sins are even worse than the physicists', because their contrafactual histories turn out after all to be all very close to some kind of a "norm," a property which my models certainly do not possess.

The foregoing argument is best illustrated by two re-interpretations of the coin-tossing record plotted on Figure 4. First of all, forgetting the origin of that figure, let us imagine that it is a geographical cross section of a new part of the world, in which all the regions below the bold horizontal lines are under water. Let us also imagine that this chart has just been brought home by an explorer and that the problem is to decide whether it was due to cause or to chance. The naïve defense will readily resort to the Highest Cause, using our graph as fresh evidence that God created Heaven and the Earth, using the same template for all the Earth, such concepts as a "continent," an "ocean," an "island," an "archipelago," or a "lake" being precisely adapted to the shape of the Earth. Against this, however, I could argue, as devil's advocate, that the Earth is a creation of blind chance, and that the possibility of using such convenient terms as "continent" and "island" just reflects the chance fact that the areas above water happen to be very short or very long very often, and to be rarely of average length.

The preceding example is not as fictitious as it may seem, because it happens precisely that the distribution of the sizes of actual islands is Paretian.[16] Hence, our hypothetical debate emphasizes the two extreme outlooks realistically, even though—the Earth having been presumably entirely explored—no actual prediction is involved in the choice between the interpretation of archipelagoes as "real" or as creations of the mind of the weary mariner.

Another example, also chosen for its lack of *direct* economic interpretation, is the problem of clusters of errors on telephone circuits. Suppose that a telephone line is only used to transmit either dots or dashes, which may be distorted in transmission to the point of being mistaken for each other. It is clear—again according to the defender of a search for causes—that whenever an electrician touches the line, one should expect to observe a small cluster of such errors. Since moreover a screwdriver touches the line many times during a single repair job, one should expect to see clusters of clusters of errors, and even clusters of third order and higher.

Actual records of the moments when errors occurred do indeed exhibit such clusters, with long periods of flawless transmission in between. A good idea of the distribution of the errors is, for example, provided by the sequence of points where the twice-used graph of Figure 4 crosses the line that earlier represented sea level. According to the searcher for causes, the precise study of such past

[16] See my "Statistics of Natural Resources and the Law of Pareto," a privately circulated "Research Note" to appear.

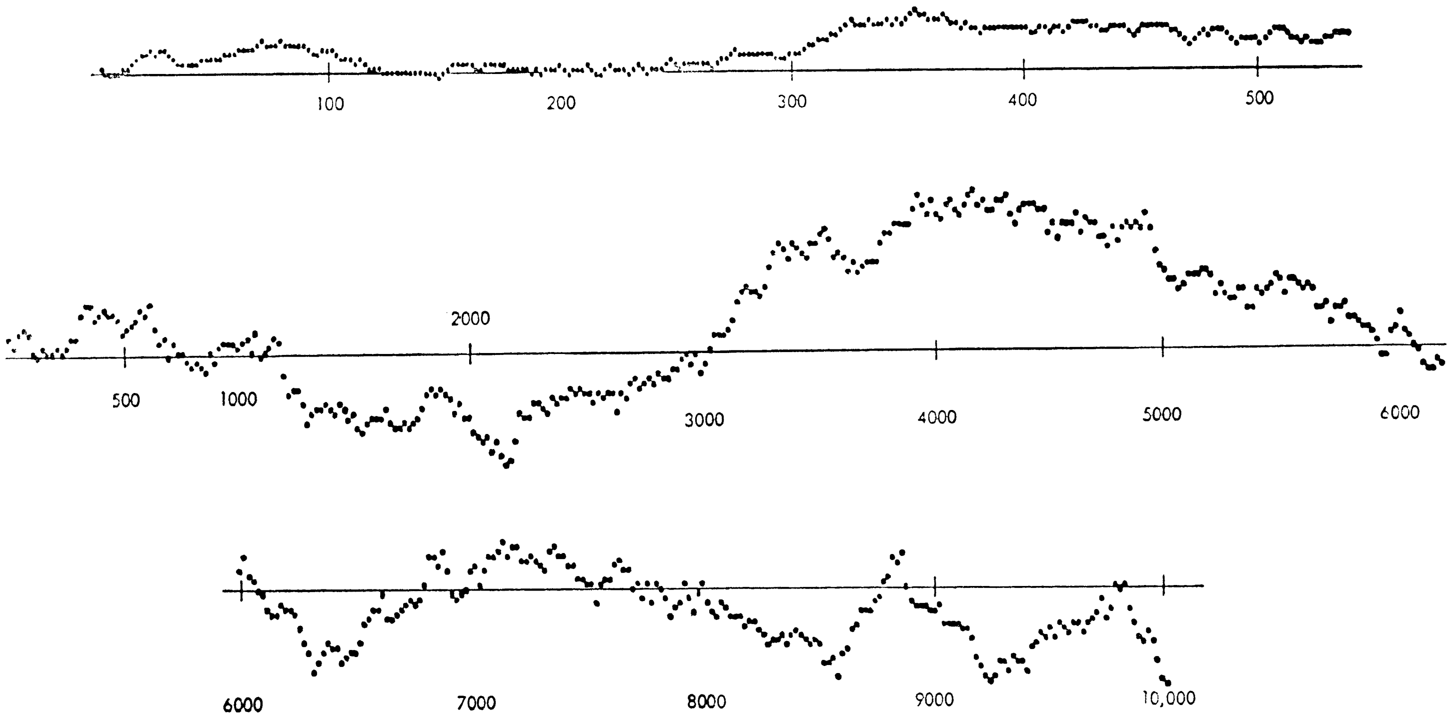

Fig. 4.—Record of Paul's winnings in a coin-tossing game, played with a fair coin. Zero-crossings seem to be strongly clustered, although intervals between crossings are obviously statistically independent. This figure is reproduced from W. Feller, *An Introduction to Probability Theory and Its Applications* (2d ed.; New York: John Wiley & Sons, 1957).

To appreciate fully the extent of apparent clustering in this figure, note that the units of time used on the second and third lines equal 20 plays. Hence, the second and third lines lack detail and each of the corresponding zero-crossings is actually a cluster, or a cluster of clusters. For example, the details of the clusters around time 200 can be clearly read on line 1, which uses a unit of time equal to 2.

records will make it possible to predict better where new errors will occur and to minimize their effects. On the other hand, precisely because of the origin of Figure 4, those beautiful hierarchies of degrees of clustering can very well be due to a "mere chance" devoid of any memory.[17]

Similar critical roles can very well be played in many contexts, and I consider it mandatory for someone to play them in relation to every important problem, without forgetting that the devil's advocate must always be on the side of the angels. An interesting example of stable truce between structure and chance is provided by the study of language and of discourse, where the traditional kind of structure is represented by grammar and—as one should expect by now—the chance mechanism is akin to the law of Pareto.[18]

VII. TWO-TAILED PARETIAN VARIABLES AND MULTIDIMENSIONAL STABLE PARETIAN LAWS

We have up to now followed tradition by associating the law of Pareto with essentially positive random variables, the distribution of which has a single long tail so that its central portion is necessarily quite skew. However, I have discovered important examples in economics of distributions having *two* Paretian tails (the most striking example refers to relative changes in the prices of sensitive speculative commodities). The argument of invariance under maximization cannot be extended to that case. But invariance under mixture simply leads to the combination of a Paretian distribution applying to all positive u and of another applying to all negative u. As to invariance under aggregation, it is satisfied by every one of the random variables constructed as either sums of or differences between two arbitrarily weighted "positive" stable variables of the kind studied earlier in this paper. In particular, these general stable variables can be symmetric; the Cauchy distribution provides a prime example. But their study depends very little upon the actual degree of skewness; hence, the asymmetry of the usual Paretian variables is less crucial than the length of their single tail.

Another remarkable property of the stable distributions is that, like the Gaussian, they have intrinsic extensions to the multivariate case, other than the degenerate case of independent co-ordinates. Very few other distributions (if any) share this property, and the reason for this is intimately related to the role of stable distributions in linear models. It is indeed possible to characterize the multivariate stable distributions as being those for which the distribution of every linear combination of the co-ordinates is a scalar stable variable. This property is essential in the study of multidimensional economic quantities, as well as in the investigation of the dependence between successive vales of a one-dimensional quantity such as income.[19]

[17] Jay M. Berger and B. Mandelbrot, "A New Model for the Clustering of Errors on Telephone Lines," *IBM Journal of Research and Development*, VII (July, 1963).

[18] See my "Linguistique statistique macroscopique," in L. Apostel, B. Mandelbrot, and A. Morf, *Logique, language et théorie de l'information* (Paris: Presses Universitaires de France, 1957), pp. 1–80 (a second edition is in preparation), and my "On the Theory of Word Frequencies and on Related Markovian Models of Discourse," in Roman Jakobson (ed.), *Structure of Language and Its Mathematical Aspects* (American Mathematical Society, 1960), pp. 190–219.

[19] See my "Stable Paretian Random Functions and the Multiplicative Variation of Income," *Econometrica*, XXIX (October, 1961), 517–43.

VIII. THE ROLE OF PARETO'S LAW IN ECONOMICS AND ESTABLISHMENT OF A LINK WITH THE PHYSICAL SCIENCES

The arguments of this paper show that there is strong pragmatic reason to begin the study of economic distributions and time series by those that satisfy the law of Pareto. Since this category includes prices,[20] firm sizes,[21] and incomes,[22] the study of Paretian laws is of fundamental importance in economic statistics.

Similarly, the example of the distribution of city sizes stresses the importance of the law of Pareto in sociology.[23] Finally, one has strong indications of its importance in psychology, but I shall not even attempt to outbid George Kingsley Zipf in listing all the Paretian phenomena of which I am aware; their number seems to increase all the time.

However, it is impossible to postpone "explanation" forever. If indeed a grand economic system is only based upon aggregation, choice, and mixture, one can prove that in order that it be Paretian, it *must* be triggered somewhere by essentially Paretian "initial" conditions. That is, however useful the method of invariants may be, it is true that it somewhat begs the question and that the basic mystery cannot be solved by pushing it around. Indeed, if it were true, in accordance with "conventional wisdom," that physical phenomena are characterized by the law of Gauss, and social phenomena by that of Pareto, we may eventually have to explain the latter by some of the "microscopic" economic models, such as the "principle" of random proportionate effect[24] which I prefer not to emphasize in my approach.

I claim, however, that such *need not* be the case. Quite the contrary, the physical world is full of Paretian phenomena that one can easily visualize as playing the role of the "triggers" that cause the economic system to be also Paretian. I have found for example[25] that single-tailed Paretian distributions, with trustworthy values for α, represent the statistical distributions of a variety of mineral resources, which are surely not influenced by the structure of society: Such is the case of the areas of oil fields and their total capacities (that is, the sums of the total production and of the currently estimated capacity); the same is true for the valuations of certain gold, uranium, and pyrite mines in South Africa. Similar findings hold for a host of similar data related to weather, which is barely influenced by man as yet—some of which data, such as hail, have a direct influence on important risk phenomena, namely insurance against hail damage, while others, such as total annual rainfall, obviously influence the sizes of crops, and hence, by the laws of supply and demand, influence the changes of agricultural prices.

If our purpose were to contribute to "geo-statistics," I should, of course, ex-

[20] See my "The Variation of Certain Speculative Prices," *Journal of Business*, Vol. XXXVI (October, 1963).

[21] See my "Oligopoly, Merger and the Paretian Size Distribution of Firms," *loc. cit.* (n. 9).

[22] See my "The Pareto-Lévy Law and the Distribution of Income," *loc. cit.* (n. 8), as amended in "The Stable Income Distribution, When the Apparent Exponent Is Near Two," *ibid.*, IV (January, 1963), 111–15; and also my "Stable Paretian Random Functions and the Multiplicative Variation of Income," *loc. cit.* (n. 19); and my "Paretian Distributions and Income Maximizations," *Quarterly Journal of Economics*, LXXVI (February, 1962), 57–85.

[23] See my "Very Long-tailed Probability Laws and the Empirical Distribution of City Sizes," in *Mathematical Explorations in Behavioral Science*, ed. F. Massarik and P. Ratoosh (Homewood, Ill.: Richard D. Irwin, Inc., in press).

[24] See my "Survey of the Growth and Diffusion Models of the Law of Pareto," *loc. cit.* (n. 5).

[25] See my "Statistics of Natural Resources and the Law of Pareto," *loc. cit.*, (n. 16).

amine the degree of generality of my claim. But, for the purpose of a study of economic time series, it will be quite sufficient to note that a Paretian grand economic system *can very well* be triggered by statistical features of the physical world. For example, natural resources and weather influence prices, which in turn influence incomes. (Since the systems to which we refer are spatio-temporal, there is nothing disturbing in our association of economic *time* series with geological and geographical *spatial* distributions.)

I shall not attempt to say anything about the actual triggering mechanism since I doubt that a unique link can be found between the social and the physical worlds. After all, quite divergent values of Pareto's α are encountered in both so that the overall grand system cannot possibly be based only upon transformations by linear aggregation, choice, and mixture.

I wish finally to point out that the Paretian phenomena of physics have also turned out to include some that are devoid of direct relation with economics. I mentioned for example that a three-dimensional stable law occurs in the theory of Newtonian attraction (see Sec. III). Moreover, the distribution of the energies of the primary cosmic rays has long been known to follow a law that happens to be identical to that of Pareto with the exponent 1.8 (as a matter of fact, Enrico Fermi's study of this problem includes an unlikely but rather neat generation for the Pareto distribution).[26] The same holds for meteorite energies, an important fact for ionospheric scatter telecommunications. Also, as discussed in Section VI, the intervals between successive errors of transmission on telephone circuits happen to be Paretian with a very small exponent, the value of which depends upon the physical properties of the circuit.

There are many reasons for believing that many Paretian phenomena are related to "accumulative" processes similar to those encountered in coin-tossing.

IX. FREDERICK MACAULAY'S CRITICISM OF THE LAW OF PARETO

Since I have found so many reasons for considering the law of Pareto to be one of the most important of all probability distributions, I have been continually surprised by the attitude to which I referred in the first sentence of this paper. I eventually realized that it had deep roots not only in the apparent lack of theoretical motivation for that law but also in several seemingly "definitive" criticisms, such as that of F. R. Macaulay.[27]

Macaulay's essay is most impressive indeed and—even though I disagree with its conclusion—I strongly recommend it because it has fully disposed of any possible claims concerning the invariance of Pareto's exponent from year to year and from country to country, and concerning the relevance of the law of Pareto to the description of small incomes or of the incomes of the lower-paid professional categories. Macaulay is also very convincing concerning Paretian distributions with a high exponent (see Sec. V).

I believe, however, that his strictures against what is called "mere curve-fitting" have been very harmful. Indeed, his ideals of a proper mathematical description are so severe that he rejects the

[26] See my "Survey of the Growth and Diffusion Models of the Law of Pareto," *loc. cit.* (n. 5.)

[27] F. Macaulay, "Pareto's Laws and the General Problem of Mathematically Describing the Frequency Distribution of Income," in *Income in the United States, Its Amount and Distribution, 1909–1919*, Vol. II (New York: National Bureau of Economic Research, 1922), chap. xxiii.

law of Pareto outright because the sample empirical curves do not "zigzag" around the simple Paretian interpolate, but rather cross it systematically a few times. This illustrates a basic difference between the outlooks of careful economists and of often careless physicists: When the law of Boyle was similarly found to differ from facts, the physicists simply invented the concept of a "perfect gas," that is, a body that follows *perfectly* Boyle's law. Naturally, perfect-gas approximations are absurd in some problems, but are adequate in many others, and they are so simple that one must consider them first. Similarly, Pareto-law approximations should not even be considered in some problems (for example, those relating to low incomes); but one must consider them first in other investigations.

Macaulay's criticism of the law of Pareto may therefore be summarized from my viewpoint by saying that it only indorses the "weak" forms of this law with which we had occasion to work. In many cases, however, I think that it is legitimate to take more seriously certain Paretian kins, such as the stable distributions.